PTO/SB/21 (09-04)
Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission

Application Number	09/710,633
Filing Date	November 8, 2000
First Named Inventor	Kent
Art Unit	1654
Examiner Name	J. E. Russel
Attorney Docket Number	TSRI 478.0 C1

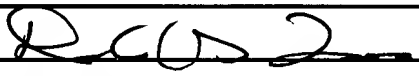
ENCLOSURES *(Check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☐ Amendment/Reply
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Reply to Missing Parts/ Incomplete Application
  - ☐ Reply to Missing Parts under 37 CFR 1.52 or 1.53
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s) _____
  - ☐ Landscape Table on CD
- ☐ After Allowance Communication to TC
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☑ Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- ☐ Proprietary Information
- ☐ Status Letter
- ☑ Other Enclosure(s) (please Identify below):
  - Postcard

Remarks

The Director is hereby authorized to charge our Deposit Account No. 19-0962 in the event there are any charges associated with the present Response or any Response in connection with this application.

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	The Scripps Research Institute		
Signature			
Printed name	Donald G. Lewis		
Date	May 16, 2007	Reg. No.	28,636

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being facsimile transmitted to the USPTO or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below:

Signature			
Typed or printed name	Donald G. Lewis	Date	May 16, 2007

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BOARD OF PATENT APPEALS AND INTERFERENCES

In re:

Applicant: Kent, et al.	)	
Serial No.: 09/710,633	)	Group Art Unit: 1654
Filed: November 8, 2000	)	Examiner: Russel, J.
Title: SYNTHESIS OF PROTEINS BY NATIVE CHEMICAL LIGATION	)	Our Ref.: TSRI 478.0Con1

Reply to Examiner's Answer

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This communication is a Reply to Examiner's Answer, mail dated 03/16/07.

Attached hereto is an Evidence Appendix containing the following references:

1. Dawson, P. E. et al., *Science* (1994), vol. 266, 776,779;
2. Ueda, H., et al., *J. Biol. Chem.* (1997), vol. 272, 24966-24960;
3. Wilken, J. et al., *Curr. Opin. Biotech.* (1998), vol. 9, 412-426; and
4. Dawson et al., *Annu. Rev. Biochem.* (2000), vol. *69*, 923-960.

Also attached hereto is an Appendix of Related Proceedings providing the following cited cases:

1. *Fromson v. Advance Offset Plate, Inc.*, 219 U.S.P.Q. 1137, 1140-1141 (Fed. Cir. 1983)
2. *In re Rasmussen*, 211 U.S.P.Q. 323, 326-27. (CCPA 1981)

Issues:

The Examiner has combined Issues No. 1-2 and 5 and maintained these issues as a single issue.

The Examiner has maintained Issue No. 3 as a separate issue.

The Examiner has withdrawn Issue No. 4.

Reply to Examiner's Answer regarding Issues 1-2 and 5:

Extrinsic Evidence of the Understanding of Persons Skilled in the Art:

The Examiner alleges that the description of the present application discloses only a single example to support claims 11-14 and 32 that this example is insufficient to reasonably convey to one skilled in the art that the inventors had possession of their invention.

The present application is based upon work that was subsequently published by the inventors in the scientific literature, viz., "Synthesis of Proteins by Native Chemical Ligation," Dawson, P. E. et al., *Science* (1994), vol 266, pages 776,779. Like the specification of the present invention, the Dawson reference discloses only a single example of the use of the claimed ligation method for synthesizing a protein having a variant amino acid, viz., [ala^{33}]IL-8 (see: page 78, Fig. 3, legend for part "A" and footnote 28). The only explanation in the Dawson reference for its choice of a mutant peptide is "convenience," see footnote 28.

The Dawson reference has been cited repeatedly by references that disclose the use of the process of claims 11-14 and 32 for making mutant proteins. These references provide extrinsic evidence that a disclosure of a scope similar to the disclosure of the present application is sufficient to teach persons skilled in the art that the ligation process of claims 11-14 and 32 can be employed for synthesizing a desired protein or domain thereof wherein the desired protein or domain thereof is "a derivative of a naturally isolatable protein that contains one or more variant residues that are not found in said naturally isolatable protein."

For example, Ueda, H., et al. (*J. Biol. Chem.* **1997**, *272*, 24966-24960) discloses a mutant cytokine (SDF-1α analogue) and cites the Dawson reference (page 24066, second column, second paragraph, footnote (15)) as disclosing the chemical ligation method employed for synthesizing this mutant cytokine. Subsequently, Wilken, J. et al., (*Curr. Opin. Biotech.* 1998, *9*, 412-426) published a detailed review of the use of this technology, including its use for making a large number of different classes proteins, including mutant proteins, see Table 2 page 416 and Figure 4, page 417. An even more extensive review was later provided by Dawson et al., *Annu. Rev. Biochem.* 2000, *69*, 923-960, see, in particular, pages 333-335 and Table 2 on page 941.

Taken together, the above references provide extrinsic evidence that a disclosure of similar scope to the description of the present application is sufficient to impart to persons skilled in the art that the inventors had possession of the invention of claims 11-14 and 32.

Judicial Support:

The courts have held that a single example can be sufficient to support a genus. *In re Rasmussen*, 650 F.2d at 1214, 211 U.S.P.Q. at 326-27, held that a disclosure of a single example of a method of "adhereingly applying" one layer to another was

sufficient to support a generic claim to "adhereingly applying" because one skilled in the art reading the specification would understand that it is unimportant how the layers are adhered, so long as they are adhered.

Claim 11 of the present application is directed to a method for ligating two peptides to form a derivative of a naturally isolatable protein having one or more **variant residues** not found in said naturally isolatable protein. Claim 32 of the present application is more narrow than Claim 11 and is directed to a method for ligating two peptides to form a desired protein or domain thereof that is a derivative of a naturally isolatable protein having one or more **cysteine residues** not found in said naturally isolatable protein. Although claims 11 and 32 differ slightly in scope, the application of *Rasmussen* to each of claims 11 and 32 is the same. However, the *Rasmussen* analysis of both claims must be divided into two instances, viz., Instance #1, wherein the variant residues or cysteines are not at the point of ligation; and Instance #2, wherein the variant residues or cysteines are at the point of ligation.

With respect to Instance #1, if the variant residues or cysteines are **not** at the point of ligation, one skilled in the art reading the specification would understand that it is unimportant with respect to the claimed ligation process whether or not variant residues or cysteines were introduced at points other than the point of ligation. A person skilled in the are would realize that the ligation process of claims 11 and 32 would proceed equally well whether nor not variant amino acids or cysteines existed at points in the resultant protein other than the point of ligation. Accordingly, according to the rule of *Rasmussen*, a specification having only one example is sufficient to support the patentability of a generic claim with respect to ligation processes falling under Instance #1.

In Instance #2, the variant residues or cysteines are at the point of ligation. It should be noted that both claims 11 and 32 impose a number of limitations on the

amino acids at the point of ligation. Both Claims 11 and 32 require that the first oligopeptide include a C-terminal thioester and that the second oligopeptide include an N-terminus having an amino acid residue with an unoxidized sulfhydryl side chain and a free amino group capable of forming a β-aminothioester likage with the C-terminal thioester that rearrages to form an amide bond therein between. Hence, claims 11 and 32 include both structural and functional limitations on the amino acid residues at the point of ligation. One skilled in the art, reading the specification, would understand that it is important for the operability of the ligation process to conform with these structural and functional limitations with respect to the amino acids at the point of ligation. However, given these limitations, one skilled in the art would also realize that it would be unimportant with respect to the operability of the claimed ligation process whether or not the amino acids at the point of ligation were variant residues not found in any naturally isolatable protein. Accordingly, in Instance #2, the introduction of variant amino acids must conform with and be consistent with other structural and functional limitations within claims 11 and 32 relating to the amino acids at the point of ligation. But given these limitations, the introduction of variants (in conformance with the limitations) would be understood by a skilled person to be unimportant to the operability of the claimed ligation process. By application of the rule of *Rasmussen*, the provision of one example in the specification is sufficient to support claims such as claims 11 and 32 with respect to Instance #2.

Examiner's Mischaracterization of Problem:

The Examiner states that the example provided by the specification, viz., Example 4, pages 37-38 and Scheme 9, page 39, "does not provide support for the genus of protein derivatives recited in claims 11-14 and 32." (Examiner's Answer, page 4, bottom paragraph) The Examiner has mischaracterized claims 11-14 and 32. Claims 11-14 are dependent claims that merely limit the products that can be produced, i.e., limitations are placed upon the type of "desired protein or domain thereof"

produced by the claimed process. The process of claims 11-14 is unaltered by any of these limitations.

Accordingly, the Examiner mischaracterizes Claims 11-14 when he states that they are directed to a "genus of protein derivatives." It would be more accurate to state that Claims 11-14 are all directed to the same process, with limitations on the product produced by said process.

Placing limitation on the product does not alter the chemistry of the claimed ligation process. The kernel of the invention lies with the ligation process steps and not with particular products produced. The products produced by the process of Claims 11-14 and 32 are mere variants and are supported by the specification.

Indeed, the variant products of claims 11-14 and 32 should not be considered limitations of the invention.

> "In general, a preamble limits the [claimed] invention if it recites essential structure or steps, or if it is 'necessary to give life, meaning, and vitality' to the claim." *Fromson v. Advance Offset Plate, Inc.*, 219 U.S.P.Q. 1137, 1140-1141 (Fed. Cir. 1983)

A person skilled in the art would appreciate that the preamble limitations of claims 11-14 and 32 do not "recite essential structure or steps," and are not "'necessary to give life, meaning, and vitality' to the claim[s]." The Examiner complains that the specification does not elaborate the significance of Example 4. However, a person skilled in the art would not require an extensive explanation of Example 4 to appreciate that the inventors possessed an understanding that the claimed process could be employed to produce variant products without variation of the claimed process.

Reply to Examiner's Answer regarding Issue 3:

The Examiner states:

> "Clearly, in view of Appellants' disclosure, a cysteine residue is required to be present at the N-terminus of one of the oligopeptides being ligated in order for the reaction to proceed. However, Example 4 and Figure 9 never discuss why a cysteine residue is present at position 41."
> (Examiner's Answer, bottom of page 6 and top of page 9)

Example 4 (page 37) discloses that the peptide corresponding to Sequence No. 12, which includes the cysteine residue at position 41, undergoes a native chemical ligation reaction with the peptide corresponding to Sequence No. 11, to yield the product HIV-1 K41, corresponding to Sequence No. 15. The disclosure of Example 4 is sufficient to impart to a person of ordinary skill in the art that, in the context of the present application, the cysteine residue present at position 41 was a necessary functionality for the native chemical ligation reaction to proceed.

Summary of Reply:

The Board is requested to reverse the Examiner's final rejection of claims 11-14 and 32 with respect to all issues.

Respectfully submitted,

Donald G. Lewis
Reg. No. 28,636
The Scripps Research Institute
10550 N. Torrey Pines Road TPC-8
San Diego, CA 92037
May 16, 2007
(858) 784-2937

Table of Contents for the Appendix of Related Proceedings:

Court of Appeals, Federal Circuit

Fromson
v. Advance Offset Plate, Inc./Graphcoat, Inc./News Publishing Company of Framingham/Newspapers of New England, Inc.

Nos. 83-850, 83-913, 83-914, 83-915

Decided Nov. 8, 1983

PATENTS

1. Infringement — In general (§39.01)

Infringement — Law or fact question (§39.60)

Infringement issue raises at least questions of what is patented and whether what is patented had been made, used, or sold by another; first is question of law, and second a question of fact.

2. Court of Appeals for the Federal Circuit — Weight given decision reviewed (§26.59)

District court's decision on infringement issue that turns on question of what is patented is reviewed as matter of law.

3. Construction of specification and claims — By specification and drawings — In general (§22.251)

Patentee's verbal license augments difficulty of understanding claims, and to understand their meaning, they are construed in connection with other parts of patent instrument and with circumstances surrounding patent application's inception.

4. Specification — Theory or principle of invention (§62.9)

Inventor need not comprehend scientific principles on which practical effectiveness of his invention rests.

5. Construction of specification and claims — Comparison with other claims (§22.40)

Significant evidence of particular claim's scope can be found on review of other claims.

6. Claims — Article defined by process of manufacture (§20.15)

Fact that process limitation appears in claim does not convert it to product by process claim.

7. Patentability — Aggregation or combination — Of old elements (§51.159)

There is no basis for treating combinations of old elements differently in determining patentability.

8. Patentability — Invention — In general (§51.501)

Analysis under 35 USC 103 for any claimed invention requires legal determination of whether claimed invention as whole would have been obvious to one of ordinary skill in art at time it was made.

9. Construction of specification and claims — By Patent Office proceedings (§22.151)

Prosecution history, or file wrapper estoppel doctrine is irrelevant when there is direct or contributory literal infringement.

10. Construction of specification and claims — By prior art (§22.20)

Claims are normally construed as they would be by those of ordinary skill in art.

Particular patents — Printing plates

3,181,461, Fromson, judgment holding claims 1, 4, 6, 7, 12, and 16 vacated.

Appeal from District Court for the District of Massachusetts, Freedman, J.; 219 USPQ 83.

Consolidated actions by Howard A. Fromson, against Advance Offset Plate, Inc.; Graphcoat, Inc., News Publishing Company of Framingham, and Newspapers of New England, Inc., for patent infringement, in which defendant counterclaims. From judgment of noninfringement, plaintiff appeals. Vacated and remanded.

John E. Lynch, New York, N.Y., for appellant.

Arthur F. Dionne, Windsor, Conn., for appellee.

Before Markey, Chief Judge, Davis, Circuit Judge, and Nichols, Senior Circuit Judge.*

Markey, Chief Judge.

Appeal from four judgments of the U.S. District Court for the District of Massachu-

* Judge Nichols assumed senior status October 1, 1983.

setts holding that claims 1, 4, 6, 7, 12 and 16 of U.S. Patent 3,181,461 issued to Fromson are not infringed or contributorily infringed. We vacate and remand.

Background

A. *The Technology*

The Fromson patent involves a process for making a photographic printing plate for use in the art of lithography. The art involves creation on a printing surface of certain areas that are hydrophilic (water attracting) and organophobic (ink repelling) and other areas that are organophilic (ink attracting) and hydrophobic (water repelling).

At the time of the Fromson invention, the state of the art was depicted generally by U.S. Patent 2,714,006, issued on July 26, 1955 to Jewett and Case (Jewett). Jewett teaches the preparation of a presensitized lithographic plate by: first treating the surface of an aluminum sheet with an aqueous solution of an alkali metal silicate to form a water insoluble, hydrophilic, siliceous, organophobic surface layer; treating that layer with a diazo compound to form a light-sensitive, water soluble, diazo coating; and exposing portions of the coated plate to light through a negative or stencil, thus causing the exposed portions to become water insoluble, hydrophobic, organophilic image areas. The plate is then washed with water to remove the water soluble diazo portions that were not exposed to light, thereby exposing the water insoluble, hydrophilic, organophobic, siliceous surfaces in their place (non-image areas). An image developer or printer's developing ink is poured on the plate and the excess wiped off, making the image areas plainly visible. The plate is then ready for mounting on a press, successive treatments with water and ink, and printing. In this process, the image areas absorb ink while the non-image areas repel it.

B. *The Fromson Patent*

In the 1950's, Fromson was in the business of selling metals and began, through Ano-Coil Corporation, to manufacture and sell anodized aluminum. In anodization, aluminum is coated with oxide while it is the anode in an electrolytic bath wherein it is subjected to an electric current, whence the term "anodized." The anodized aluminum was used in articles such as television antennas, furniture tubing, and nameplates.

Fromson, with no background in lithography, conceived of using anodized aluminum as a replacement for non-anodized aluminum in the plate taught by Jewett. His invention according to the Fromson patent improves the Jewett plate in a number of ways. It enables use in preparation of the plate of light-sensitive compounds other than diazo compounds. It enables the coating to absorb nitrogen-containing materials released by the light-sensitive compounds when exposed to light. Also, as the district court found, the Fromson plate enjoys improved corrosion resistance and a longer press life.

Fromson filed his application for patent in May 1963, and the patent issued in May 1965, containing eleven product and five process claims. Claim 1 is representative of the product claims:

> 1. A sensitized photographic printing plate comprising an aluminum sheet having a surface which has been treated to form an aluminum oxide coating on said surface, a water-insoluble, hydrophilic, organophobic layer on said sheet resulting from the reaction of the aluminum oxide coating and an alkali metal silicate applied to said coating, and a light-sensitive coating over said layer [e.g., diazo resin] having one solubility in relation to a solvent in a state before exposure to light and another solubility in relation to said solvent in another state after exposure to light, said light-sensitive material being soluble in said solvent in one of said states and being insoluble in said solvent and in water, hydrophobic and organophilic in its other state.

Claim 12 is the sole independent process claim:

> 12. The process of making a sensitized photographic plate comprising the steps of applying to an aluminum sheet having a coating of aluminum oxide, a water-solution of an alkali metal silicate to cause the silicate to react with the aluminum oxide to form a water-insoluble, hydrophilic, organophobic layer on said sheet, drying the layer, and applying over the dry layer a light-sensitive coating having one solubility in relation to a solvent in a state before exposure to light and another solubility in relation to said solvent in another state after exposure to light, said light-sensitive material being soluble in said solvent in one of said states and being insoluble in said solvent and in water, hydrophobic and organophilic in its other state.

C. *The Advance Plate*

After issuance of his patent, Fromson's invention enjoyed extensive commercial success and was the subject of licensing agreements with several companies. Fromson sued

Advance Offset Plate, Inc. ("Advance"), charging it with infringement and contributory infringement of product claims 1, 4, 6 and 7, and process claims 12 and 16. He also sued three Advance customers for direct infringement in actions consolidated with that against Advance and now consolidated on appeal. We assume for purposes of this appeal that the claims at issue in those three cases are the same as those asserted against Advance.

It is undisputed that Advance manufactured and sold "wipe-on" plates comprising an anodized aluminum sheet that had been treated with an aqueous solution of sodium silicate, and that its customers applied a diazo coating to those plates. Because the claims include the application of a diazo coating or other light sensitive layer and because Advance's customers, not Advance, applied the diazo coating, Advance cannot be liable for direct infringement with respect to those plates but could be liable for contributory infringement. It is also undisputed that Advance, not its customers, applied the diazo resin to certain "presensitized" plates. For those presensitized plates, Advance could be liable for direct infringement.

Though preparation of the Advance plate involves treatment of anodized aluminum with an aqueous solution of alkali metal silicate to yield a water-insoluble, hydrophilic layer, as set forth in the claims, Advance and its customers deny infringement on the sole ground that there is no "reaction" between the aluminum oxide and sodium silicate. It is undisputed that Advance and its customers do what the claims say, i.e., apply a water solution of an alkali metal silicate to an oxide coated aluminum sheet to produce a layer, but it is argued that the layer does not result from a "reaction" as we are asked by Advance to define that term.

During oral argument on appeal, Advance suggested that its process involves different conditions (e.g., temperature, time) than the Fromson process, an argument neither raised in the brief nor addressed by the district court. Though the argument cannot therefore be considered here, we caution that the asserted claims contain no temperature or time limitations, and that no basis appears on this record for limiting the claimed inventions to preferred embodiments or specific examples in the specification. See, Smith v. Snow, 294 U.S. 1, 11, 24 USPQ 26, 36 (1935).

D. District Court's Decision

By Memorandum and Order dated May 30, 1980, the district court separated the issues of infringement and invalidity from other defenses and counterclaims. On February 11, 1983, after trial without a jury, the district court found no infringement of "the patent." Because the finding could apply properly only to the six claims at issue, we review the decision as if only the six claims were involved. ——— F. Supp. ———, 219 USPQ 83 (D. Mass. 1983). The district court did not resolve the validity issue, noting that the patent had expired. If infringement during the period before expiration be found, the validity issue would of course become viable.

The district court held that the claims of the Fromson patent, when read in light of the specification and prosecution history, must be interpreted as requiring that the water-insoluble, hydrophilic, organophobic layer be the product of a chemical reaction between the aluminum oxide coating and the alkali metal silicate, and that that product be a compound having physical properties different from those of its constituents. Finding no such reaction in preparation of Advance's plates, the district court entered a judgment of noninfringement in the four consolidated actions.

The district court considered unpersuasive three laboratory tests and testimony of Fromson's expert, based on those tests, that Advance's plate had a surface layer resulting from reaction of sodium silicate and anodized aluminum. It considered persuasive the tests and testimony of Advance's expert, who concluded that there is no reaction product of aluminum oxide and silicate on Advance's plate, but that the surface is coated with silica, an insoluble, hydrophilic material. Resolution of that conflicting testimony is not necessary, however, where, as here, a legal conclusion establishes its irrelevance. Indeed, we accept for this opinion that the Advance plate involves the formation of silica.

The district court made no finding on whether silica is organophobic. In response to request for admission number 29, Advance stated that it employs a sodium silicate coating believed hydrophilic and organophobic. The district court found that silica (not sodium silicate) is formed on Advance's plate. On remand, a finding of infringement will require a determination of whether the layer on Advance's plates is water insoluble, hydrophilic *and* organophobic.

The district court made no finding on the identity of the layer produced by the process described in the Fromson patent as the preferred process embodiment (i.e., 155-210°F. and 1-10 minutes of treatment time). At trial, Advance's expert said that the Fromson process does not produce aluminosilicate within those time and temperature parameters. He

also said, however, that although silica was first formed when he treated aluminum oxide crystals with an aqueous solution of an alkali metal silicate within those parameters, aluminosilicate could form after treatment lasting one hour.

E. Reissue Proceeding

On March 29, 1979, after the four suits were filed, Fromson filed for reissue in the Patent and Trademark Office ("PTO") under a procedure which then allowed for reissue proceedings and reexamination in light of new prior art. Advance participated as a protestor in accordance with PTO rules. Before the judgment in the district court was rendered, the PTO again concluded that Fromson's claims were patentable.

Issue

Whether the district court erred in finding no infringement or contributory infringement of claims 1, 4, 6, 7, 12, and 16.

Opinion

[1,2] The issue of infringement raises at least two questions: (1) what is patented, and (2) has what is patented been made, used or sold by another. SSIH Equipment S.A. v. USITC, 713 F.2d 746, 758, 218 USPQ 678, 688 (Fed. Cir. 1983). The first is a question of law; the second a question of fact. Id.; Kalman v. Kimberly-Clark Corp., 713 F.2d 760, 771, 218 USPQ 781, 788, 789 (Fed. Cir. 1983). The present decision of the district court turned on the first question, which we review as a matter of law.

A. Contentions of the Parties

Advance and its customers contend that "the Court was totally convinced, based on the evidence presented, that Fromson was indeed referring to and claiming a 'reaction product' formed by the reaction of an aluminum oxide with sodium silicate, i.e., an aluminosilicate compound." We agree that the district court so interpreted the claims.

Fromson argues that "reaction" in the claims should be interpreted to cover the claimed treatment of an oxide coated aluminum sheet with an aqueous solution of alkali metal silicate to form a water insoluble, hydrophilic, organophobic layer on the sheet, and that whether the layer is an aluminosilicate compound is irrelevant, there being no reference to any such compound in the asserted claims. We agree.

B. Claim Construction — The Specification

[3] In Autogiro Co. of America v. United States, 384 F.2d 391, 397, 155 USPQ 697, 702 (Ct. Cl. 1967), our predecessor court recognized that patentees are not confined to normal dictionary meanings:

> The dictionary does not always keep abreast of the inventor. It cannot. Things are not made for the sake of words but words for things. To overcome this lag, patent law allows the inventor to be his own lexicographer. (Citations omitted.)

A patentee's verbal license "augments the difficulty of understanding the claims," and to understand their meaning, they must be construed "in connection with the other parts of the patent instrument and with the circumstances surrounding the inception of the patent application." Id. Accord, General Electric Co. v. United States, 572 F.2d 745, 751-53, 198 USPQ 65, 70-73 (Ct. Cl. 1978).

This appeal hinges on construction of "reaction." The specification discloses a new and improved method of forming plates for use in lithography. Fromson discovered that the treatment of anodized aluminum with an aqueous solution of water soluble alkali metal silicate produces a water insoluble, hydrophilic, organophobic layer on the aluminum, a layer having exceptional lithography-related properties. Fromson's invention included the formation of the layer, not its exact structure. Though Fromson referred to the disclosed treatment as involving a "reaction," he also referred to it in the specification as an "application" and as "adsorption." Not all references to "reaction" were accompanied by a reference to formation of an aluminosilicate.

[4] Fromson did theorize that his new, improved layer was an aluminosilicate believed "to be in the nature of a commercial zeolite," having "properties of a molecular sieve," but expressed that theory as merely a "belief." There is no basis or warrant for incorporating that belief as a limitation in the claims. It is undisputed that inclusion of Fromson's theory and belief was unnecessary to meet the enablement requirement of 35 U.S.C. §112 (that a patentee describe how to make and use the invention). Moreover, it is axiomatic that an inventor need not comprehend the scientific principles on which the practical effectiveness of his invention rests. See, e.g., Diamond Rubber Co. v. Consolidated Rubber Co., 220 U.S. 428, 435-36 (1911).

C. Claim Construction — The Claims

[5] Significant evidence of the scope of a particular claim can be found on review of

other claims. General Electric v. United States, supra, 572 F.2d at 752, 198 USPQ at 70. Here, claim 5 (not asserted) limits the layer described in claim 1 to "an aluminosilicate structure in the nature of a zeolite molecular sieve", i.e., to Fromson's theory of what is formed. In Kalman v. Kimberly-Clark Corp., supra, 713 F.2d at 771, 218 USPQ at 788, this court said "where some claims are broad and others narrow, the narrow claim limitations cannot be read into the broad whether to avoid invalidity or to escape infringement." Accord, Environmental Designs, Ltd. v. Union Oil Co. of California, 713 F.2d 693, 700, 218 USPQ 865, 871 (Fed. Cir. 1983); Caterpillar Tractor Co. v. Berco, S.P.A., 714 F.2d 1110, _____, 219 USPQ 185, 188 (Fed. Cir. 1983). The aluminosilicate limitation of narrow claim 5 cannot, therefore, be read into broader claim 1.

[6,7,8] Advance and its customers argue that, had appellant originally sought product by process claims, those claims would have been disallowed because each claimed element would have been old, citing the district court's finding 18 containing "combination of old element" terminology. This argument is without merit. First, it is a hypothetical. That a process limitation appears in a claim does not convert it to a product by process claim. Second, there is no basis for treating combinations of old elements differently in determining patentability. The analysis under 35 U.S.C. §103 for any claimed invention requires a legal determination of whether the claimed invention as a whole would have been obvious to one of ordinary skill in the art at the time it was made. See, e.g., Environmental Designs, Ltd. v. Union Oil of California, supra; Stratoflex, Inc. v. Aeroquip Corp., 713 F.2d 1530, 218 USPQ 271 (Fed. Cir. 1983).

D. Claim Construction — Prosecution History

The district court noted some of Fromson's arguments during prosecution of his application, in which he stressed the importance of "reacting" anodized aluminum with alkali metal silicate. However, whether the interaction of these two materials was a "reaction" or something else was immaterial to consideration of the prior art. It does not appear, moreover, that Fromson used "reacting" in his arguments any differently than he had in the specification and claims, i.e., to describe what he believed the interaction was between oxide coated aluminum and an aqueous solution of alkali metal silicate. Thus, Fromson's arguments focused on the fact of an interaction and production of a new layer with particular properties, not on the specific nature of the interaction or on any chemical structure of the layer.

That Fromson speculated, on one page of a response to a rejection, that the reaction layer is "believed to be in the nature of a commercial zeolite" is of no moment, in view of the total absence from the other thirteen pages in that response of any reference to formation of an aluminosilicate or zeolite, and in view of his clear labeling of the zeolite statement as a "belief." Instead, Fromson referred in those thirteen pages to an aluminum oxide-sodium silicate reaction surface or layer, indicating that he did not know, and did not care, what the "reaction" or the structure of the resulting product might be.

[9] Advance and its customers argue that the "prime issue" in this appeal is the doctrine of prosecution history ("file wrapper") estoppel. That doctrine, however, is inapplicable here. If there be literal infringement, direct and contributory (as there may be here under a proper construction of the claims), the doctrine is irrelevant. Even if it were applicable, an examination of a prosecution history demonstrating that "reaction" was merely a theoretical label having no influence on the patentability of the claimed invention as a whole demonstrates the absence of prosecution history estoppel.

E. Claim Construction — Other Considerations

That "reaction" in the claims need not be confined to production of an aluminosilicate is consistent with the dictionary definition. That in Webster's New Collegiate Dictionary (1974) includes both "chemical transformation or change," and "interaction of chemical entities," which are consistent with the definitions appearing in Hackh's Chemical Dictionary (1969) and the American Heritage Dictionary (1970).

When an oxide coated aluminum surface is contacted with an aqueous solution of water soluble alkali metal silicate, chemical change occurs in at least two ways. First, ions or other chemical units in solution have somehow interacted to form a solid structure. Second, the water insoluble solid structure, whatever may be its precise nature (e.g., silica or aluminosilicate), is not identical to the water alkali metal silicate and oxidized aluminum that interacted to produce it. Moreover, there is clearly present an "interaction of chemical entities."

[10] The foregoing is fully consistent with long-standing use of "reaction" in the lithog-

raphy art. Claims are normally construed as they would be by those of ordinary skill in the art. See e.g., Schenck v. Nortron Corp., 713 F.2d 789, 786-787, 218 USPQ 698, 701-02 (Fed. Cir. 1983). Jewett interchangeably uses terms such as "treating," "treatment," and "react," to describe a lithographic plate producing process. Jewett's claims use "reacting," "treatment," and "reaction product." Jewett makes no attempt to define the structure of the layer there disclosed (as an aluminosilicate compound or otherwise), although it does mention the hydrophilic layer as being chemically bonded to the aluminum surface. Jewett refers to the layer as "silicate treatment," as "silicate or silicon containing" film, or as "an inorganic material such as silicate." It is not unreasonable to conclude that one of ordinary skill in the lithography art would interpret "react" in Fromson to mean the same thing it appears to mean in Jewett, i.e., the treatment of a metal substrate with an aqueous solution to yield a layer, regardless of the chemical structure of the layer or the proper label for the phenomena that produced it.

Conclusion

We hold, therefore, that the district court erred as a matter of law in interpreting the claims as limited to the product of a chemical reaction producing a new chemical compound in the restrictive sense of those terms. There having been no finding by the district court that the accused plates have a layer that is water insoluble, hydrophilic *and* organophobic, the determination of infringement and contributory infringement must await complete findings in the first instance by the district court.

Decision

The four judgments based on findings of noninfringement of Fromson's asserted claims are vacated, and the cases are remanded for further consideration consistent with this opinion.

Vacated and remanded.

Court of Appeals, Federal Circuit

The Young Engineers, Inc. (aka TYE or TYE, Inc.),
v. United States International Trade Commission

No. 83-649
Decided Nov. 8, 1983

UNFAIR COMPETITION

1. Importation restrained under Tariff Act (§68.60)

Exclusion order and cease and desist order are alternative remedies.

2. Importation restrained under Tariff Act (§68.60)

There is no authority in 19 USC 1337(h) to support action of ITC that changed its orders after twelve-month period.

3. Importation restrained under Tariff Act (§69.60)

While Commission determinations are not final for purposes of appeal to CAFC until review period has run, they are otherwise effective upon publication in Federal Register, 19 USC 1337(g)(2); during Presidential review period, products are in fact excluded from entry except under bond; thus, Commission orders are "effective" before they are "final," not "final" and then "effective."

4. Importation restrained under Tariff Act (§68.60)

19 USC 1337(h) relates to order in effect and how long it will "continue in effect"; moreover, this section recognizes that order disapproved by President under Section 1337(g) is no longer in effect at earlier date; there is no authority in Section 1337(h) to issue exclusion orders but only to end effectiveness of outstanding order where conditions that led to such exclusion from entry or order no longer exist; 5 CFR 211.57 does not provide greater authority than statute.

5. Importation restrained under Tariff Act (§68.60)

"Determination" in 19 USC 1337(b) means only conclusion that there is violation or no violation of Section 1337(a); although word "determination" appears in other paragraphs of Section 1337, its meaning is confused rather than clarified by comparison of various clauses in which it appears.

6. Importation restrained under Tariff Act (§68.60)

19 USC 1337(b)(1) time limitations are satisfied if Commission completes its investi-

utile per inutile non-vitialtur might appropriately be applied here.

Court where the verbiage of synergism is used.[1] The maxim utile per inutile non vitiatur might appropriately be applied here.

Moreover, I should have been content to continue the present ambiguous stance with respect to synergism described in note 17 of the majority opinion, rather than to stand up and be counted in the current synergism controversy.[2]

Perhaps the word synergism should be discarded (except for its original meaning with respect to the interaction of chemicals or drugs, and for its use as a fashionable fad in television commercials, and for its theological and scriptural overtones).[3] But abandoning the verbal trappings and "rhetoric of synergism" must not cause courts to overlook the importance of the requirement of novelty and invention, long required by the patent statutes and the Constitution.

As pointed out in John Deere itself, Congress may not "enlarge the patent monopoly without regard to the innovation, advancement or social benefit gained thereby. Moreover, Congress may not authorize the issuance of patents whose effects are to remove existent knowledge from the public domain, or to restrict free access to materials already available. Innovation, advancement, and things which add to the sum of useful knowledge are inherent requisites in a patent system which by constitutional command must 'promote the Progress of * * * useful Arts.' This is the *standard* expressed in the Constitution and it may not be ignored." 383 U.S. at 6, 148 USPQ at 462.

Similarly, this Court has said: "Thus, the courts, in determining obviousness in a combination patent, must undertake the tripartite Graham inquiry without losing sight of the necessity to determine whether the device performs its function in an innovative fashion." 608 F.2d at 91, 203 USPQ at 965-966.

It must never be forgotten that the power given to Congress by Art. I, sec. 8, cl. 8 of the Constitution is "*To promote the Progress of Science* and useful Arts, by securing for limited Times to Authors and *Inventors* the exclusive Right to their respective Writings and *Discoveries*." [Italics supplied] The primary policy of the patent laws is to promote invention for the benefit of the public. The private gain enjoyed by the patentee is secondary; the "exclusive Right" conferred by the patent monopoly is merely the means of accomplishing the intended result of advancing the growth of science by adding to the sum of human knowledge. A patent cannot be sustained which would withdraw or subtract from what is already known and practiced. Borden Co. v. Clearfield Cheese Co., 244 F.Supp. 366, 368, 146 USPQ 660, 661 (W.D. Pa. 1965). To fence in by a newly created monopoly elements previously available to the public (by aggregating them in a combination patent without any inventive innovation) would be contrary to public policy and fundamental principles of patent law.

To emphasize the importance of these constitutional aspects of our patent system, whether or not they are clothed in "the rhetoric of synergism," it seemed proper to dwell upon them specifically in this concurring opinion when joining in the judgment of the Court.

Court of Customs and Patent Appeals

In re Rasmussen

No. 81-516

Decided June 4, 1981

PATENTS

1. Amendments to patent application — New matter (§13.5)

Claims — Specification must support (§20.85)

35 U.S.C. 132 prohibits introduction of new matter into disclosure of application; 35 U.S.C. 112, first paragraph, requires that claim language be supported in specification.

2. Amendments to patent application — New matter (§13.5)

Claims — Broad or narrow — In general (§20.201)

Specification — Claims as disclosure (§62.3)

Broadening claim does not add new matter to disclosure; disclosure is that which is taught, not that which is claimed; original claim is part of disclosure at time of

filing; consideration of original claim as evidencing support in disclosure for later submitted claims does not warrant employment of 35 U.S.C. 132 as basis for rejection of later submitted claims on ground that latter are adding new matter to original claim portion of disclosure; to do so would render Section 132 redundant in light of Section 112, first paragraph; applicant is entitled to claims as broad as prior art and his disclosure will allow.

3. Amendments to patent application — New matter (§13.5)

Claims — Specification must support (§20.85)

Pleading and practice in Patent Office — Rejections (§54.7)

Proper basis for rejection of claim amended to recite elements thought to be without support in original disclosure is 35 U.S.C. 112, first paragraph, not Section 132; latter section prohibits addition of new matter to original disclosure; it is properly employed as basis for objection to amendments to abstract, specifications, or drawings attempting to add new disclosures to that originally presented; past opinions of Court of Customs and Patent Appeals, in cases in which Section 132 claim rejection was reviewed on Section 112 analysis, should not in future be viewed as having approved employment of Section 132 as basis for claim rejection; amended claims involved in those cases should have been rejected under Section 112, first paragraph; claim rejections in those cases could then have been explicitly affirmed or reversed on direct applications of Section 112, rather than on Section 112 analyses applied to Section 132 rejections; similarly, rejections of claims for lack of support when required in reissue applications should be made under Section 112, first paragraph, rather than under new matter prohibition of 35 U.S.C. 251; accordingly, such cases are overruled insofar as they approved rejection of claims under Section 132.

4. Claims — Broad or narrow — In general (§20.201)

Reissue — In general (§58.1)

Fact that claim may be broader than specific embodiment disclosed in specification is in itself of no moment; statutory provision for broadened claims in reissue applications is intended to meet precisely situation in which patentee has claimed less than he had right to claim.

5. Pleading and practice in Patent Office — In general (§54.1)

Specification — Sufficiency of disclosure (§62.7)

35 U.S.C. 112 requires disclosure of only one mode of practicing invention; insistence upon boilerplate recitation in specification that specific embodiment shown was not meant to limit breadth of claims, or that example given was only one of several methods that could be employed, is exaltation of form over substance.

Appeal from Patent and Trademark Office Board of Appeals.

Application for reissue of patent of Max Otto Henri Rasmussen, Serial No. 884,775, filed Mar. 8, 1978, for reissue of Patent No. 3,963,549, issued June 15, 1976. From decision affirming rejection of claim 6, applicant appeals. Reversed; Nies, J., dissenting.

George Vande Sande, Washington, D.C., for appellant.

Joseph F. Nakamura (Robert D. Edmonds, of counsel) for Patent and Trademark Office.

Before Markey, Chief Judge, and Rich, Baldwin, Miller, and Nies, Associate Judges.

Markey, Chief Judge.

The decision of the Patent and Trademark Office Board of Appeals (board) affirming the rejection of claim 6 under 35 USC 132 is reversed.

Background

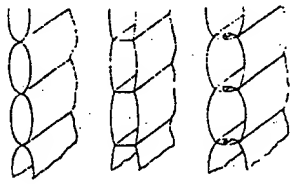
Appealed claim 6 is contained in reissue application S.N. 884,775 filed March 8, 1978.[1] Original claim 6 in the reissue application was directed to a method of manufacturing a thermal insulating member.[2] Rasmussen described in his specification the steps of applying adhesive to one side of a tubular plastic film, winding the film around two spaced drums, and, when the desired number of layers have been wound, cutting the film layers

---

[1] That application seeks reissue of U.S. Patent 3,963,549, issued June 15, 1976.

[2] 6. A method of manufacturing a thermal insulating member from a thin film of plastic material and comprising a pair of spaced opposing generally parallel sidewalls which are bridged by a plurality of spaced transverse walls comprising the steps of:

transversely to the direction of winding. When the film units are unfolded and extended, a plastic laminate in one of the forms shown below results:



Viewing the particular method of adhering the layers of tubular film as immaterial, Rasmussen later amended claim 6, inter alia, by substituting "adheringly applying" for language specifying use of adhesives.[3]

The examiner rejected the amended claim, saying "adheringly applying" was "new matter" prohibited by §132, explaining that limitation of the scope of the original disclosure to use of adhesives meant that allowance of the broader claim would be an enlargement of the scope of the disclosure.

The board affirmed, saying Rasmussen's application disclosed only one embodiment (applying adhesive to join the sheets) and that broadening the scope of the claim added new matter to the application.

Issue

The issue presented is whether amended claim 6 was properly rejected under 35 USC 132.

Opinion

[1] Confusion is generated when related but distinct statutory provisions are treated as interchangeable. Section 132 prohibits the introduction of new matter into the disclosure of an application. Section 112, first paragraph, requires that claim language be supported in the specification. This court, having said that a rejection of an amended claim under §132 is equivalent to a rejection under §112, first paragraph, for lack of support, appears to have contributed to the treatment of those separate statutory sections as interchangeable. See In re Hogan, 559 F.2d 595, 608, 194 USPQ 527, 539 (CC-

winding a continuous length of a tube formed of the plastic material in its flattened state into a generally cylindrical member whose circumference corresponds to the desired width of the insulating member,

applying to the flattened tube a band of adhesive of predetermined width corresponding substantially to the width of said spaced transverse walls and thus to the desired spacing between the sidewalls, said application of adhesive occurring prior to the contacting of the flattened tube during the winding step to the portion of the tube already wound onto the cylindrical member so that successive layers of said tubular member on said cylindrical member adhere to each other along the predetermined width of adhesive application,

terminating the winding of the plastic tube onto the cylindrical member when a predetermined number of layers of the plastic tube has been wound thereon corresponding to the desired length of the thermal insulating member,

and cutting the superimposed assemblage of successively adhering layers of the plastic tube in a direction transverse to the longitudinal direction of the tube,

said assemblage when longitudinally straightened and then extended in a direction transverse to the longitudinal direction of the tube so as to expand the successively joined tubes forming said thermal insulating member. [Emphasis added.]

[3] Amended claim 6 reads as follows:

6. A method of manufacturing a thermal insulating member from a thin film of material and comprising a pair of spaced opposing generally parallel sidewalls which are bridged by a plurality of spaced transverse walls comprising the steps of:

winding a continuous length of a tube formed of the material in its flattened state onto a generally cylindrical member whose circumference corresponds to the desired width of the insulating member, each successive layer of said tube being wound to overlie the immediately preceding layer to provide thereby a generally cylindrical band of tubular layers extending axially along said cylindrical member a distance corresponding substantially to the flattened width of said tube,

adheringly applying the flattened tube during the winding step to the portion of the tube already wound onto the cylindrical member *over a band of predetermined width* corresponding substantially to the desired width of said spaced transverse walls and thus to the desired spacing between said sidewalls,

terminating the winding of the tube onto the cylindrical member when a predetermined number of layers of the tube has been wound thereon corresponding to the desired length of the thermal insulating member,

and cutting the superimposed assemblage of successively adhering layers of the tube in a direction transverse to the longitudinal direction of the tube,

said assemblage when longitudinally straightened and then extended in a direction transverse to the longitudinal direction of the tube so as to expand the successively joined tubes forming said thermal insulating member. [Emphasis added.]

PA 1977), In re Wertheim, 541 F.2d 257, 265, 191 USPQ 90, 99 (CCPA 1976), In re Bowen, 492 F.2d 859, 864, 181 USPQ 48, 52 (CCPA 1974), In re Smyth, 480 F.2d 1376, 1385, 178 USPQ 279, 286 (CCPA 1973).[4]

Apparently reluctant to reverse on the sole ground that an improper statutory provision had been employed, and recognizing the burden on the parties inherent in a return of the case for application of §112, this court has reviewed §132 claim rejections on the basis of whether the rejected claim found support in the original disclosure. See In re Eickmeyer, 602 F.2d 974, 981, 202 USPQ 655, 662 (CCPA 1979); In re Barker, 559 F.2d 588, 593-94, 194 USPQ 470, 474 (CCPA 1977); In re Winkhous, 527 F.2d 637, 640, 188 USPQ 129, 131 (CCPA 1975).

Similarly, new matter rejections of claims under §251 have been reviewed on the basis of a §112 analysis, that is, on whether a claim found support in an original patent. See In re East, 495 F.2d 1361, 1366, 181 USPQ 716, 719 (CCPA 1974).

As is illustrated in the present case, employment of §§132 and 112 as interchangeable leads to confusion of two distinct concepts: (1) the adding of new matter to the disclosure, and (2) the broadening of a claim.

[2] Broadening a claim does not add new matter to the disclosure. Disclosure is that which is taught, not that which is claimed.[5] An applicant is entitled to claims as broad as the prior art and his disclosure will allow.

[3] The proper basis for rejection of a claim amended to recite elements thought to be without support in the original disclosure, therefore, is §112, first paragraph, not §132. The latter section prohibits addition of new matter to the original disclosure. It is properly employed as a basis for objection to amendments to the abstract, specifications, or drawings attempting to add new disclosure to that originally presented. Past opinions of this court, in cases in which a §132 claim rejection was reviewed on a §112 analysis, should not in future be viewed as having approved the employment of §132 as a basis for claim rejection. The amended claims involved in those cases should have been rejected under §112, first paragraph. The claim rejections in those cases could then have been explicitly affirmed or reversed on direct applications of §112, rather than on §112 analyses applied to §132 rejections.[6] Accordingly, such cases are overruled insofar as they approved rejection of claims under §132.

Turning to the merits of this appeal, we will again treat a §132 claim rejection before us as though it had been made under §112, first paragraph. We proceed to decide the case on that basis in the interest of judicial economy. Were we to merely reverse the rejection as having been made under an inappropriate statutory provision, and say no more, the PTO would presumably enter a rejection under §112 and that decision would then be appealable to this court.

Amended claim 6 recites the adhering step as "adheringly applying" one layer of tube to an adjacent earlier layer. Rasmussen's specification describes that step as follows: "[A]dhesive is applied to the tubular foil 4 in a narrow or broader strip, possibly in two narrow strips. Accordingly, the face of the tubular foil successively sticks to the winding lying on the drums." The language of the specification thus describes one method of "adheringly applying" one layer to the other.

[4] As above indicated, that a claim may be broader than the specific embodiment disclosed in a specification is in itself of no moment. Indeed, the statutory provision for broadened claims in reissue applications is intended to meet precisely the situation in which a patentee has claimed "less" than he had a right to claim. 35 USC 251.

In In re Smythe, 480 F.2d 1376, 1384, 178 USPQ 279, 285 (CCPA 1973), this court stated:

> A hypothetical situation may make our point clear. If the original specification of

[4] MPEP 706.03(o), 608.04-608.04(c), and 1411.02 relate to considerations set forth herein.

[5] We deal here with rejection of amended claims, and, by implication, with rejection of entire new claims submitted after filing. An original claim is part of the disclosure at the time of filing. In re Anderson, 471 F.2d 1237, 1238, 176 USPQ 331, 332 (CCPA 1973). Consideration of an original claim as evidencing support in the disclosure for later submitted claims does not warrant employment of §132 as a basis for rejection of later submitted claims on the ground that the latter are adding new matter to the original claim portion of the disclosure. To so hold would render §132 redundant in light of §112, first paragraph.

[6] Similarly, rejections of *claims* for lack of support when required in reissue applications should be made under §112, first paragraph, rather than under the new matter prohibition of 35 USC 251.

a patent application on the scales of justice disclosed only a 1-pound "*lead* weight" as a counterbalance to determine the weight of a pound of flesh, we do not believe the applicant should be prevented, by the so-called "description requirement" of the first paragraph of §112, or the prohibition against new matter of §132, from later claiming the counterbalance as a "metal weight" or simply as a 1-pound "weight" although both "metal weight" and "weight" would indeed be progressively broader than "lead weight," including even such an undisclosed, but obviously art-recognized equivalent, "weight" as a pound of feathers. The broader claim language would be permitted because the *description of the use and function* of the lead weight as a scale counterbalance in the *whole disclosure* would immediately convey to any person skilled in the scale art the knowledge that the applicant invented a scale with a 1-pound counterbalance weight, regardless of its composition.

[5] Similarly, one skilled in the art who read Rasmussen's specification would understand that it is unimportant *how* the layers are adhered, so long as they are adhered.[1] Thus the phrase "adheringly applying" is supported by the example found in the specification.

Conclusion

The phrase "adheringly applying" being supported in the specification, rejection of that claim under 35 USC 132, first paragraph, is *reversed*. Rejection under the appropriate statutory provision, 35 USC 112, would have been inappropriate.

Reversed

[1] The board seemed to realize that 35 USC 112 requires disclosure of only one mode of practicing the invention, but nevertheless insisted upon a boilerplate recitation in the specification that the specific embodiment shown was not meant to limit the breadth of the claims, or that the example given was only one of several methods which could be employed. Such insistence is here an exaltation of form over substance.

District Court, S.D. New York

Hedaya Brothers, Inc.
v. Capital Plastics, Inc.

No. 79 Civ. 4104
Decided July 23, 1980

COPYRIGHTS

1. Infringement — Tests of (§24.209)

Test for determining whether copyright infringement has occurred is whether average lay observer would find substantial similarity in designs, recognizing copy as appropriation of copyrighted work; plaintiff in copyright infringement action can only prevail if overall aesthetic impressions created by designs are substantially same.

2. Infringement — In general (§24.201)

Matter copyrightable — In general (§24.301)

In making determination of copyright infringement, it must be remembered that plaintiff's copyright does not protect idea of design, but only plaintiff's particular expression of that idea; even if defendant has sedulously borrowed each of plaintiff's ideas, that alone is not violative of copyright statute.

3. Infringement — In general (§24.201)

That ideas contained in design were old hat and originality of design was minimal must also be weighed in considering claim of infringement because, where basic design is not original with plaintiff small variations by subsequent designers may protect them from charges of infringement.

4. Infringement — Tests of (§24.209)

Accused has not infringed plaintiff's copyrights where aesthetic appeals of parties' respective designs are not substantially same, and would not strike average lay observer as being substantially same.

5. In general (§24.01)

Notice of copyright (§24.35)

Copyright infringement action is governed by Copyright Act of 1909 where designs in question were created and published prior to January 1, 1978, effective date of Copyright Act of 1976; under old Act, statutory protection was obtained only by publication of work with notice of copyright required by what was formerly 17 U.S.C. 10, in form required by former 17 U.S.C. 19; while under saving provision of former 17

Table of Contents for the Evidence Appendix:

data. Another method much in vogue then was the use of the analytical ultracentrifuge, by which, on the same time photograms used for measuring the sedimentation coefficient (by monitoring the movement of a boundary through Schlieren optics), one could also assess the diffusion coefficient (*18*).

More recent methods for measuring diffusion coefficients of macromolecules are based on dynamic laser light scattering (*19*). Clearly, all these methods require dedicated equipment and are not easily available to scientists in nonspecialized laboratories. The present method is based on commonly available equipment (an instrument for capillary zone electrophoresis or just a capillary, pump, and UV detector), now standard in most biochemical laboratories, and allows an easy and reproducible determination of D values for both small analytes and macromolecules. However, a relatively large difference between D_{exp} and D_{tab} for macromolecules indicates that slower flows should be used in such cases.

The present method allows for a quick and precise estimation of the molecular diffusion coefficient and, thus, of the radius of a molecule in a wide range of molecular mass values and might be useful in a larger number of chemical and biochemical laboratories than in the past. An additional advantage of this application of Taylor's approach is that only small volumes of solution (fractions of nanoliters) are required.

REFERENCES AND NOTES

1. G. Taylor, *Proc. R. Soc. London Ser. A* 219, 186 (1953).
2. G. K. Batchelor, *An Introduction to Fluid Dynamics* (Cambridge Univ. Press, Cambridge, 1970).
3. G. Taylor, *Proc. R. Soc. London Ser. A* 225, 473 (1954).
4. R. Aris, *ibid.* 235, 67 (1956). In his original paper (*1*), Taylor neglected axial molecular diffusion and obtained only the second, although the most important, term in Eq. 2. That limited application of the theory, as specified in (*3*), to the case $RU/D >> 6.9$. This limitation is not restrictive; however, Eq. 2 given by Aris is valid even for very slow flows in narrow-bore tubes where axial molecular diffusion may be comparable to Taylor's dispersion.
5. The characteristic diffusion time for a solute in a tube is, by order of magnitude, the time necessary for a solute particle to diffuse over the capillary cross section given by R^2/D. In (*1*) it was defined as the time necessary for a radial nonuniformity to reduce its amplitude e times and was found to be given by $R^2/3.8^2 D$.
6. A. Bournia, J. Coull, G. Houghton, *Proc. R. Soc. London Ser. A* 261, 227 (1961).
7. H. R. Bailey and W. B. Gogarty, *ibid.* 269, 352 (1962).
8. E. V. Evans and C. N. Kenney, *ibid.* 284, 540 (1965).
9. W. N. Gill, *ibid.* 298, 335 (1967).
10. _____ and R. Sankarasubramanian, *ibid.* 316, 341 (1970).
11. It is worth remembering that the theory and experimental methodology for capillary zone electrophoresis had been developed by the mid-1970s [S. Hjertén, *Chromatogr. Rev.* 9, 122 (1967); R. Virtanen, *Acta Polytech. Scand.* 123, 1 (1974)], although, extensive applications and further development of the method did not begin until a decade later [J. W. Jorgenson and K. D. Lukacs, *Science* 222, 266 (1983)].
12. Equation 4 is obtained by setting $k = 0$ in equations 13 and 14 of E. Grushka, in *Methods of Protein Separation*, N. Catsimpoolas, Ed. (Plenum, New York, 1975), vol. 1, p. 161.
13. This allowed for recalculation of the found coefficients to another temperature, provided the dependence of the viscosity on temperature is known. As diluted water solutions or distilled water was used, the viscosity dependence on temperature was assumed to be that of water in all the cases [*CRC Handbook of Chemistry and Physics* (CRC Press, Boca Raton, FL, 1987), p. F-37].
14. *American Institute of Physics Handbook* (McGraw-Hill, New York, 1957), pp. 2–193.
15. The capillary surface was initially coated with linear polyacrylamide, according to the method of S. Hjertén, *J. Chromatogr.* 347, 191 (1985). Then, a dextran layer was grafted to the polyacrylamide as reported by K. Ganzler *et al.*, *Anal. Chem.* 64, 2665 (1992).
16. H. R. Mahler and E. H. Cordes, *Biological Chemistry* (Harper and Row, New York, ed. 2, 1971).
17. R. H. Stokes, *J. Am. Chem. Soc.* 72, 763 (1950).
18. H. Neurath, *Chem. Rev.* 30, 357 (1942).
19. B. J. Berne and R. Pecora, *Dynamic Light Scattering* (Wiley, New York, 1976).

20. Supported in part by a grant from Consiglio Nazionale delle Ricerche, Comitati di Chimica e Medicina e Biologia, and Radius in Biotechnology (European Space Agency, Paris) to P.G.R.

27 May 1994; accepted 15 September 1994

Synthesis of Proteins by Native Chemical Ligation

Philip E. Dawson, Tom W. Muir, Iain Clark-Lewis, Stephen B. H. Kent*

A simple technique has been devised that allows the direct synthesis of native backbone proteins of moderate size. Chemoselective reaction of two unprotected peptide segments gives an initial thioester-linked species. Spontaneous rearrangement of this transient intermediate yields a full-length product with a native peptide bond at the ligation site. The utility of native chemical ligation was demonstrated by the one-step preparation of a cytokine containing multiple disulfides. The polypeptide ligation product was folded and oxidized to form the native disulfide-containing protein molecule. Native chemical ligation is an important step toward the general application of chemistry to proteins.

Proteins owe their diverse properties to the precisely folded three-dimensional structures of their polypeptide chains. This is the defining feature of a protein, rather than size or molecular mass per se. Merely describing the three-dimensional structure of a protein is insufficient to fully explain its biological properties. A better understanding of how structure dictates the biological properties of a protein would be achieved by systematically varying the covalent structure of the molecule and correlating the effects with the folded structure and biological function.

In this report, we describe an important extension of the chemical ligation method (*1*) to allow the preparation of proteins with native backbone structures. The principle of "native chemical ligation" is shown in Fig. 1. The first step is the chemoselective reaction of an unprotected synthetic peptide-α-thioester (*2*, *3*) with another unprotected peptide segment containing an amino-terminal Cys residue to give a thioester-linked intermediate as the initial covalent product. Without change in the reaction conditions, this intermediate undergoes spontaneous, rapid intramolecular reaction to form a native peptide bond at the ligation site. The target full-length polypeptide product is obtained in the desired final form without further manipulation. We believe that general synthetic access of this type will allow almost unlimited variation of the covalent structure of the protein molecule.

Model studies were undertaken with small peptides to investigate the native chemical ligation approach (*4*). These studies were consistent with the mechanism shown in Fig. 1, in which the initial thioester ligation product was not observed as a discrete intermediate because of the rapid rearrangement to form a stable peptide bond. Facile intramolecular reaction results from the favorable geometric arrangement of the α-NH_2 moiety with respect to the thioester formed in the initial chemoselective ligation reaction. The use of such "entropy activation" for peptide bond formation is based on principles enunciated by Brenner (*5*) and more recently adopted by others (*6*).

Study of a variety of model peptides established that native chemical ligation was generally applicable to peptides containing the full range of functional groups

P. E. Dawson, T. W. Muir, S. B. H. Kent, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA.
I. Clark-Lewis, Biomedical Research Center, Department of Biochemistry, 2222 Health Sciences Mall, University of British Columbia, Vancouver, BC V6T 1Z3 Canada.

*To whom correspondence should be addressed.

normally found in proteins (7). As described in this report, native chemical ligation is limited to reaction at an amino-terminal Cys residue. It was important to prevent the side chain thiol of this Cys from oxidizing to form a disulfide-linked dimer, because this was unreactive in the ligation. An excess of thiol corresponding to the thioester leaving group was used to keep the Cys residues in reduced form without interfering with the ligation reaction. The amino-terminal peptide segment must be prepared by chemical synthesis to equip it with the necessary α-COSR functionality (where R is an alkyl group) (2). Furthermore, for optimal ligation, this component should have an unhindered (that is, non β-branched) carboxyl-terminal amino acid. Solubilizing agents such as urea or guanidine hydrochloride did not interfere with the ligation and could be used to enhance the concentration of peptide segments, and thus increase the reaction rate.

Further model reactions showed that the use of better thioester leaving groups resulted in faster ligation reactions. We applied this observation to the native chemical ligation of peptides from the extracellular domain of a human cytokine receptor (8) (Fig. 2). Use of the 5-thio-2-nitrobenzoic acid (-SNB) leaving group, corresponding to the reduced form of Elman's reagent, gave rapid reaction in high yield. As described in the legend to Fig. 2, the reaction between the peptide segments was observed to have gone essentially to completion in less than 5 min, giving the 50-residue product with a native peptide bond at the site of ligation. Thus, rapid native chemical ligation can be achieved by use of a thioester leaving group with suitably tuned properties.

Application of the native chemical ligation method to the total synthesis of a protein molecule was illustrated by the preparation of human interleukin 8 (IL-8) (9). The 72–amino acid polypeptide chain contains four Cys residues, which form two functionally critical disulfide bridges in the native protein molecule (9). The total synthesis of IL-8 is shown in Fig. 3. The two unprotected synthetic peptide segments reacted cleanly to give the full-length polypeptide chain in reduced form without further chemical manipulation (10). This successful ligation was particularly significant because the 33- and 39-residue IL-8 segments each contained two Cys residues and together encompassed 18 of the 20 genetically encoded amino acids found in proteins. The purified product was folded and oxidized as previously described (9) to give IL-8 with a mass precisely 4 daltons less than that of the original ligation product, indicating the formation of two disulfide bonds. The properties of this folded product were identical to those of authentic IL-8 samples (11). This result unambiguously confirmed the formation of a peptide bond at the ligation site, because the thioester-to-amide rearrangement must have taken place to give the free Cys^{34} side chain that formed the native disulfide bond (see Fig. 3A).

What is likely to be the impact of native chemical ligation on the study of proteins? Proteins are usually studied by expression in genetically engineered microorganisms with the methods of recombinant DNA-based molecular biology. Methods such as site-directed mutagenesis (12) have had a revolutionary impact on the ability to prepare large numbers of modified proteins in useful amounts for systematic study (13). Innovative approaches have increased the range of amino acids that can be incorporated in

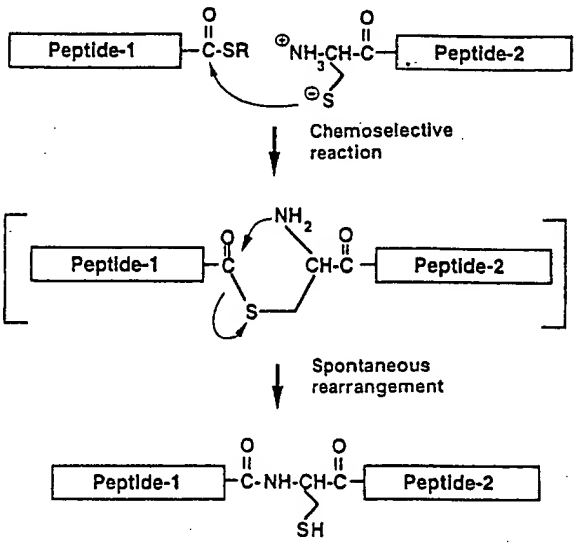


Fig. 1. The principle of native chemical ligation. The synthetic segment, peptide 1, which contains a thioester at the α-carboxyl group, undergoes nucleophilic attack by the side chain of the Cys residue at the amino terminal of peptide 2 (R is an alkyl group). The initial thioester ligation product undergoes rapid intramolecular reaction because of the favorable geometric arrangement [involving a five-membered ring] of the α-amino group of peptide 2, to yield a product with a native peptide bond at the ligation site. Both reacting peptide segments are in completely unprotected form, and the target peptide is obtained in final form without further manipulation.

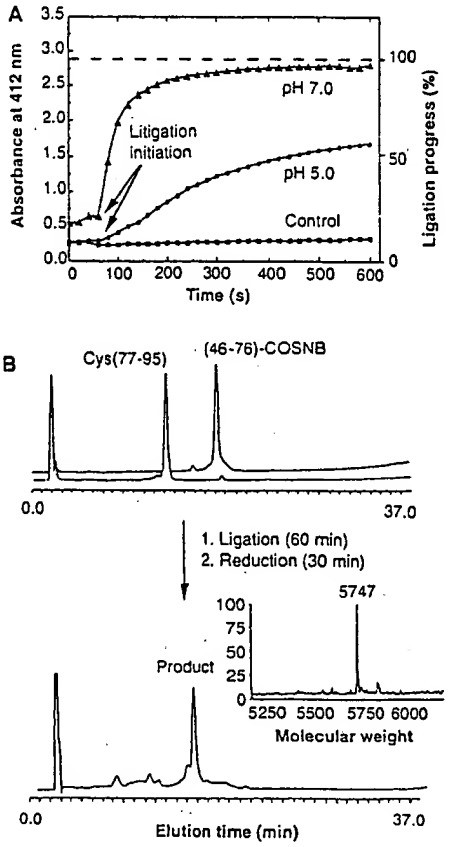


Fig. 2. Rapid native chemical ligation reaction, illustrated by the synthesis of a peptide segment corresponding to residues 46 to 95 from the external domain of the human IL-3 receptor β-subunit (8). (**A**) Monitoring by ultraviolet (UV) absorbance. Ligation was initiated by adding [Cys^{77}](77-95) to purified Msc(46-76)αCOSNB (27) at the stated pH and the reaction was monitored by UV [the substituted aryl thiolate leaving group has a characteristic UV absorbtion at 412 nm ($\varepsilon_{TNB,\ 412\ nm}$ = 13,700 $dm^3\ mol^{-1}\ cm^{-1}$)]. At pH 7.0, the reaction was essentially complete within 5 min. No reaction was observed when Msc(46-76)αCOSNB was exposed to a 10-fold molar excess of Leu-enkephalin (amino-terminal residue, Tyr) at pH 5.0. This control experiment confirms the absolute requirement for an amino-terminal Cys residue at the site of ligation. (**B**) Monitoring by HPLC. Purified [Cys^{77}](77-95) (0.98 mM) and (45-76)αCOSNB (0.9 mM) were reacted in 8 M urea, pH 5.0, 50 mM ammonium acetate buffer at 23°C. Analytical HPLC (C_{18} reversed phase; 22.5 to 45% acetonitrile at 0.7% per minute; monitored at 214 nm) of the individual components is shown (upper trace). After 1 hour, the ligation solution was exposed to the reducing agent tris(2-carboxyethyl)phosphine at pH 8.0 and subsequently raised to pH 13 to remove the Nα-Msc moiety. Analytical HPLC, under the same conditions, of the crude product is shown (lower trace). The 50-residue product had the expected molecular mass by electrospray mass spectrometry [observed, 5747.0 daltons; calculated (average isotope composition), 5747.4 daltons]. The ligation product was shown to be stable at high pH and to reducing conditions, and formed an intramolecular disulfide. These observations are consistent with the presence of a native peptide bond at the site of ligation.

expression systems (*14*) and promise to significantly extend the utility of biosynthetic modification of the covalent structure of proteins. However, there appear to be limitations inherent to the nature of ribosomal protein synthesis (*14*).

In favorable cases, chemical synthesis has already made important contributions to the exploration of the relationship of protein structure to function. Stepwise solid phase synthesis has permitted the de novo preparation of small proteins (*15*), and there have been several notable examples of the use of this method of total protein synthesis to explore the molecular basis of biological function (*16*). Another method that has in special instances allowed chemistry to be applied to the study of proteins is semisynthesis through the conformationally assisted religation of peptide fragments (*17*). An important extension of the semisynthesis approach is the use of enzymatic ligation of cloned or synthetic peptide segments (*18*). Although these methods currently have severe limitations, there continues to be serious interest in the wider application of the tools of organic chemistry to the study of proteins (*15*).

Recently, we introduced the chemical ligation of unprotected peptide segments as an improved route to the total synthesis of proteins (*1*). The key aspect of this approach was the use of chemoselective reaction to specifically and unambiguously join peptides by formation of an unnatural (that is, nonpeptide) backbone structure at the ligation site. It has permitted the facile preparation of a wide range of backbone-modified proteins, including analogs of protein domains (*19*) and of the human immunodeficiency virus–1 (HIV-1) proteolytic enzyme (*1*, *20*). Chemical ligation has also proven to be useful for the routine, reproducible synthesis of large amounts of proteins in high purity with full biological activity (*21*), as well as for the straightforward production of protein-like molecules of unusual topology (*22*). However, the range of proteins accessible by this technique is limited by the size of the synthetic peptide segments (*23*). A useful extension would occur if we had direct synthetic access to native backbone polypeptide chains up to the size of typical protein domains (*24*). Chemical ligation would then allow us to string these domains together to explore the world of proteins in a general fashion.

Native chemical ligation provides precisely that capability. It combines the formation of a native peptide bond at the ligation site with the advantages of chemoselective reaction of unprotected peptides (*1*). This second-generation ligation chemistry dramatically increases the size of native backbone polypeptides directly accessible by total chemical synthesis (*25*). It can be usefully applied to a wide range of synthetic targets, including proteins of moderate size, and it allows direct access to protein functional domains (*24*). Native chemical ligation is a foundation stone of a general modular approach to the total chemical synthesis of proteins (*26*).

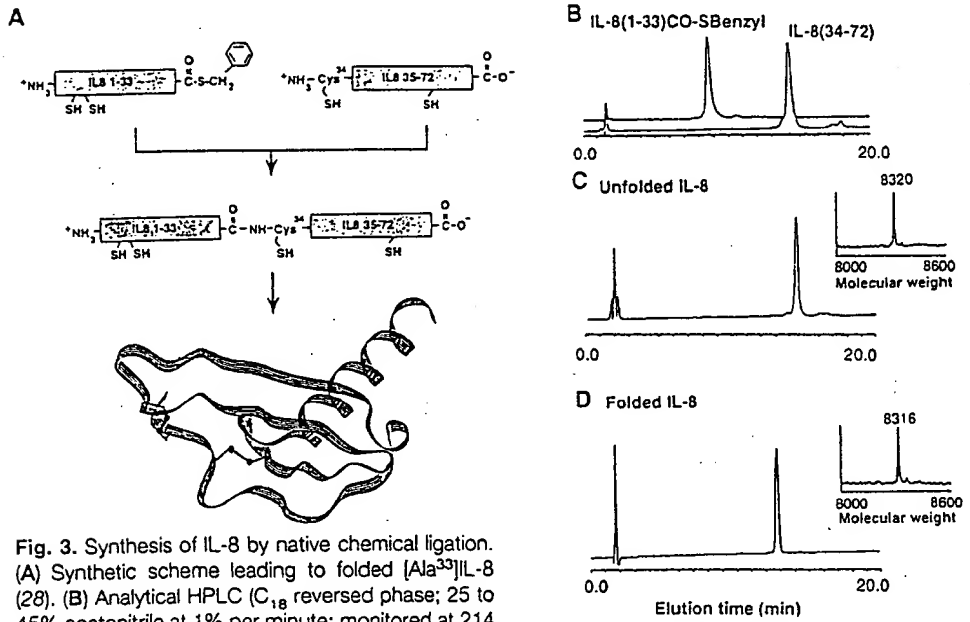


Fig. 3. Synthesis of IL-8 by native chemical ligation. (**A**) Synthetic scheme leading to folded [Ala^{33}]IL-8 (*28*). (**B**) Analytical HPLC (C_{18} reversed phase; 25 to 45% acetonitrile at 1% per minute; monitored at 214 nm) of the synthetic peptide segments (*29*), IL-8(1-33)αCOSBzl and IL-8(34-72), each shown before reaction was initiated (*30*). (**C**) Analytical HPLC under the same conditions of the purified ligation product, IL-8(1-72)(SH)$_4$, in fully reduced form. (Inset) Electrospray mass spectrum (raw data displayed as a single charge state): observed molecular mass 8319.8 daltons; calculated molecular mass (average isotope composition), 8319.8 daltons. (**D**) Air oxidation of the purified 1-72 ligation product to form the folded [Ala^{33}]IL-8 molecule, shown after HPLC purification. The earlier elution of the folded, disulfide cross-linked native protein compared with the reduced polypeptide is typical (*9*). Folding and oxidation conditions: polypeptide at 0.2 mg/ml, 1 M guanidine·HCl, pH 8.5 tris buffer, and vigorous stirring in air at ambient temperature. (Inset) Electrospray mass spectrometry of the oxidized and folded synthetic IL-8 (raw data displayed as a single charge state). Observed molecular mass, 8315.6 daltons; calculated molecular mass (average isotope composition), 8315.8 daltons.

REFERENCES AND NOTES

1. Chemical ligation [M. Schnölzer and S. B. H. Kent, *Science* 256, 221 (1992)] involved the chemoselective reaction of unprotected peptides to give a product with an unnatural backbone structure at the ligation site. Use of unprotected peptides circumvented the difficulties inherent to classical chemical synthesis, namely, complex combinations of protecting groups that lead to limited solubility of many synthetic intermediates [for example, K. Akaji *et al.*, *Chem. Pharm. Bull. (Tokyo)* 33, 184 (1985)]. In contrast, the chemical ligation technique has allowed us to make good use of the ability to routinely make, purify, and characterize unprotected peptides 50 or more residues in length (*23*).
2. Peptide 1 in Fig. 1. The key feature is the -αCOSR moiety, which is readily generated from the peptide-αCOSH prepared by highly optimized stepwise solid-phase peptide synthesis on a thioester resin. The thioester resin was prepared by a generalized version (L. E. Canne and S. B. H. Kent, manuscript in preparation; details can be obtained from the authors) of the Blake-Yamashiro procedure (*3*). Peptide products were cleaved, purified, and characterized by standard methods (*23*).
3. J. Blake, *Int. J. Pept. Protein Res.* 17, 273 (1981); D. Yamashiro and C. H. Li, *ibid.* 31, 322 (1988).
4. To help explore the mechanism of the reaction, the peptide Leu-Tyr-Arg-Ala-Gly-αCOSBzl (Bzl, benzyl) was reacted with Ac-Cys (Ac, acetyl). The exact mass of the resulting ligation product was determined by electrospray mass spectrometry and was consistent with a thioester-linked peptide as the ligation product generated by nucleophilic attack of the Ac-Cys side chain on the α-thioester moiety of the peptide. Reaction of Leu-Tyr-Arg-Ala-Gly-αCOSBzl with H-Cys-Arg-Ala-Glu-Tyr-Ser (containing an unblocked α-NH$_2$ functional group) proceeded rapidly at pH 6.8 (below pH 6, the reaction proceeded very slowly, suggesting the involvement of the ionized thiolate form of the Cys side chain) and gave a single product of the expected mass. This product lacked susceptibility to nucleophiles and had the ability to form disulfide-linked dimeric peptides, indicating unambiguously the formation of a native amide bond at the ligation site.
5. M. Brenner, in *Peptides. Proceedings of the Eighth European Peptide Symposium*, H. C. Beyerman, Ed. (North-Holland, Amsterdam, 1967), pp. 1–7.
6. D. S. Kemp and R. I. Carey, *J. Org. Chem.* 58, 2216 (1993); C.-F. Liu and J. P. Tam, *J. Am. Chem. Soc.* 116, 4149 (1994).
7. Even free internal Cys residues may be present in either of the reacting segments. Internal Cys residues can undergo ester exchange with the peptide-α-thioester component; however, this reaction is unproductive because no rearrangement to the amide bond can occur; the thioester formed is readily reversible and remains a productive part of the reacting system.
8. R. D'Andrea *et al.*, *Blood* 83, 2802 (1994).
9. M. Baggiolini and I. Clark-Lewis, *FEBS Lett.* 307, 97 (1989); I. Clark-Lewis, B. Dewald, M. Loetscher, B. Moser, M. Baggiolini, *J. Biol. Chem.* 269, 16075 (1994); I. Clark-Lewis, *Biochemistry* 30, 3128 (1991); K. Rajarathnam, I. Clark-Lewis, B. D. Sykes, *ibid.* 29, 1689 (1994).
10. Analogous methods have required removal of protecting groups (*3*, *6*) or conversion of intermediates to the final form (*6*), or both steps. No previous method has allowed the chemical reaction of unprotected peptide segments to directly yield a native backbone final product.
11. Titration in an assay for neutrophil elastase releas

(9) demonstrated that the potencies [median effective dose (ED_{50}) = 0.3 nM] and maximal responses of the folded, ligated [Ala^{33}]IL-8 and the corresponding molecule obtained by conventional synthesis (9) were indistinguishable and identical to native sequence IL-8.

12. M. Smith, *Angew. Chem. Int. Ed. Engl.* 33, 1214 (1994).
13. C. Eigenbrot and A. Kossiakoff, *Curr. Opin. Biotechnol.* 3, 333 (1992).
14. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* 244, 182 (1989); J. A. Ellman, D. Mendel, P. G. Schultz, *ibid.* 255, 197 (1992); V. W. Cornish *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 91, 2910 (1994).
15. T. W. Muir and S. B. H. Kent, *Curr. Opin. Biotechnol.* 4, 420 (1993).
16. M. Miller *et al.*, *Science* 246, 1149 (1989); A. Wlodawer *et al.*, *ibid.* 245, 616 (1989); L. H. Huang, H. Cheng, A. Pardi, J. P. Tam, W. V. Sweeney, *Biochemistry* 30, 7402 (1991); K. Rajarathnam *et al.*, *Science* 264, 90 (1994).
17. R. E. Offord, in *Protein Design and the Development of New Therapeutics and Vaccines*, J. B. Hook and G. Poste, Eds. (Plenum, New York, 1990), pp. 253–282; C. J. A. Wallace and I. Clark-Lewis, *J. Biol. Chem.* 267, 3852 (1992).
18. L. Abrahmsen *et al.*, *Biochemistry* 30, 4151 (1991); T. K. Chang, D. Y. Jackson, J. P. Burnier, J. A. Wells, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
19. Ligated 10F3, the integrin-binding module of fibronectin: 95 residues (M. Williams, T. Muir, M. Ginsberg, S. B. H. Kent, *J. Am. Chem. Soc.*, in press).
20. Catalytic contribution of flap-substrate hydrogen bonds in HIV-1 protease explored by chemical synthesis: homodimer of 99-residue subunits [M. Baca and S. B. H. Kent, *Proc. Natl. Acad. Sci. U.S.A.* 90, 11638 (1993)].
21. R. C. deLisle Milton, S. C. F. Milton, M. Schnölzer, S. B. H. Kent, in *Techniques in Protein Chemistry IV* (Academic Press, New York, 1992), pp. 257–267.
22. Four-helix bundle template-assembled synthetic protein: molecular mass 6647 daltons [P. E. Dawson and S. B. H. Kent, *J. Am. Chem. Soc.* 115, 7263 (1993)]; homogeneous multivalent artificial protein: molecular mass 19,916 daltons [K. Rose, *ibid.* 116, 30 (1994)]; artificial neoprotein mimic of the cytoplasmic domains of a multichain integrin receptor: molecular mass 14,194 daltons [T. W. Muir, M. J. Williams, M. H. Ginsberg, S. B. H. Kent, *Biochemistry* 33, 7701 (1994)]; peptide dendrimer: molecular mass 24,205 daltons [C. Rao and J. P. Tam, *J. Am. Chem. Soc.* 116, 6975 (1994)].
23. By using optimized stepwise solid-phase methods [M. Schnolzer, P. Alewood, D. Alewood, S. B. H. Kent, *Int. J. Pept. Protein Res.* 40, 180 (1992)], the preparation in good yield and high purity of peptides up to 60 residues is routine; in favorable cases, peptides with >80 residues can be prepared.
24. A. L. Berman, E. Kolker, E. N. Trifonov, *Proc. Natl. Acad. Sci. U.S.A.* 91, 4044 (1994).
25. The carboxyl-terminal peptide segment or protein module could also be expressed by standard recombinant DNA methods; provided the product contained an amino-terminal Cys residue, it could be reacted with the synthetic amino-terminal peptide-αCOSR by using the native chemical ligation described here to give a product in which part of the protein had been derived from chemical synthesis and part from ribosomal synthesis.
26. A modular strategy for the total synthesis of proteins has been developed based on the convergent chemical ligation of unprotected peptides [L. E. Canne, S. K. Burley, S. B. H. Kent, paper presented at the Annual Meeting of the Protein Society, San Diego, July 1994]. Protein domains (modules) were prepared by chemical ligation of 50 to 70 residue segments; these domains were then stitched together to give the target protein. Mutually compatible ligation chemistries are required: Intradomain ligation should optimally yield a stable, peptide-like bond, whereas interdomain ligation will tolerate a wider variation of properties of the structure formed at the ligation site.
27. Crude synthetic Msc(46-76)αCOSH [Msc, 2(methylsulfonyl)ethyloxycarbonyl] was converted to the 5-thio-2-nitrobenzoic acid ester (-COSNB) by treatment with 5,5'-dithio-bis(2-nitrobenzoic acid) [10 equivalents (eq)] in 8 M urea, pH 4.0 50 mM ammonium acetate buffer. This thioester-containing material was found to be completely stable below pH 6.0, and was readily purified by reversed-phase high-performance liquid chromatography (HPLC).
28. [Ala^{33}]IL-8 was chosen as the synthetic target for convenience; previous work had shown that this mutant IL-8 had full biological activity (9), and a supply of the Boc-Ala (Boc, butyloxycarbonyl) thioester resin was on hand for other applications. The folded structure shown is based on the x-ray structure of the IL-8 monomer [E. T. Baldwin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 88, 502 (1991)].
29. The IL-8 peptide segments were prepared by optimized stepwise solid-phase synthesis (23) and were purified by reversed-phase HPLC and characterized by standard methods. Crude synthetic segment IL-8(1-33)αCOSH was converted to the thiobenzyl ester by reaction with benzyl bromide (15 eq) in 6 M guanidine-HCl, pH 4.6, sodium acetate buffer, prior to purification under standard reversed-phase HPLC conditions.
30. The segments (1-33)αCOSBzl (5.0 mg, 1.3 μmol) and 34-72 (4.8 mg, 1.1 μmol) were reacted in 0.5 ml 6 M guanidine-HCl, pH 7.6, phosphate buffer at 23°C in the presence of benzyl mercaptan (5 μl)]. After suitable reaction time (48 to 72 hours), a ligation yield of ~60% was obtained. The product was purified by reversed-phase HPLC and characterized by electrospray mass spectrometry.
31. We gratefully acknowledge the assistance of T. Walters and M. Baca in the early stages of this work, of L. Canne in providing the thioester resin, of B. Dewald for some of the elastase release assays, and of R. Simon and S. Clark for critical comments on the manuscript. Supported by funding from NIH [GM48897-01 and GM48870-03 (S.B.K.); GM 50969-01 (I.C.L.)].

16 August 1994; accepted 21 September 1994

A Theropod Dinosaur Embryo and the Affinities of the Flaming Cliffs Dinosaur Eggs

Mark A. Norell,* James M. Clark,† Dashzeveg Demberelyin, Barsbold Rhinchen, Luis M. Chiappe, Amy R. Davidson, Malcolm C. McKenna, Perle Altangerel, Michael J. Novacek

An embryonic skeleton of a nonavian theropod dinosaur was found preserved in an egg from Upper Cretaceous rocks in the Gobi Desert of Mongolia. Cranial features identify the embryo as a member of Oviraptoridae. Two embryo-sized skulls of dromaeosaurids, similar to that of *Velociraptor*, were also recovered in the nest. The eggshell microstructure is similar to that of ratite birds and is of a type common in the Djadokhta Formation at the Flaming Cliffs (Bayn Dzak). Discovery of a nest of such eggs at the Flaming Cliffs in 1923, beneath the *Oviraptor philoceratops* holotype, suggests that this dinosaur may have been a brooding adult.

Dinosaur eggs are abundant in Upper Cretaceous rocks of the Gobi Desert (*1*, *2*), but embryonic skeletons from these deposits are scarce. Definitive remains include numerous bird embryos (*3*) and a single fragmentary specimen of an ornithischian hind limb (*4*). Because the definitive taxonomic identity of eggs requires the presence of identifiable embryonic remains within them, the identity of most egg types present in Upper Cretaceous beds in Mongolia has been unclear (*1*, *2*).

In 1993, a rich Upper Cretaceous fossil locality in the Gobi Desert was discovered (*5*). The site, Ukhaa Tolgod, is in the northeastern Nemegt Basin, Omnogov Aimak, near the salt extraction settlement of Daus. In addition to over 300 mammal and lizard skulls, 20 theropod skeletons (including several adult and juvenile oviraptorids), and many protoceratopsian and ankylosaurid dinosaurs discovered at this locality, at least five types of eggs were found. Many of these were arranged in nests. One egg, from a heavily weathered nest, contains the nearly complete skeleton of an embryonic oviraptorid dinosaur (Fig. 1). Also among the broken eggshell fragments in this nest were two tiny skulls (~5 cm long) of a dromaeosaurid theropod, one preserved with eggshell adhering to it (Fig. 2).

The red sandstones of Ukhaa Tolgod probably belong to either the Djadokhta Formation or the Barun Goyot Formation and lie 35 km east of the Barun Goyot type section (*6*). Limited studies indicate the presence of taxa typical of either or both the Barun Goyot and the Djadokhta formations (that is, *Velociraptor*, *Mononykus*, and the mammals *Zalambdalestes*, *Bulgaanbaatar*, *Nemegtbaatar*, and *Catopsalis*). These faunas are considered correlative with the Judithian North American land mammal age and the Campanian marine stage (*7*), although this correlation is poorly constrained (*8*).

M. A. Norell, J. M. Clark, L. M. Chiappe, A. R. Davidson, M. C. McKenna, M. J. Novacek, Department of Vertebrate Paleontology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA.
Dashzeveg D. and Barsbold R., Geological Institute, Mongolian Academy of Sciences, Ulaan Baatar 11, Mongolia.
Perle A., Mongolian Museum of Natural History, Ulaan Baatar 46, Zaluchudiin Gudarna-1, Mongolia.

*To whom correspondence should be addressed.
†Present address: Department of Biological Sciences, George Washington University, Washington, DC 20052, USA.

THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 272, No. 40, Issue of October 3, pp. 24966–24970, 1997
Printed in U.S.A.

Chemically Synthesized SDF-1α Analogue, N33A, Is a Potent Chemotactic Agent for CXCR4/Fusin/LESTR-expressing Human Leukocytes*

(Received for publication, June 26, 1997, and in revised form, July 30, 1997)

Hirotsugu Ueda, Michael A. Siani‡, Wanghua Gong§, Darren A. Thompson‡, Garth G. Brown¶, and Ji Ming Wang‖

From the Laboratory of Molecular Immunoregulation, Division of Basic Sciences and §Intramural Research Support Program, SAIC Fredrick, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, Maryland 21702, ¶NEN™ Life Science Products, Iodination Group, Billerica, Massachusetts 01862, and ‡Gryphon Sciences, South San Francisco, California 94080

Stromal cell-derived factor (SDF) 1 is a potent chemoattractant for leukocytes through activation of the receptor CXCR4/Fusin/LESTR, which is a fusion co-factor for the entry of T lymphocytotropic human immunodeficiency virus type 1 (HIV-1). This CXCR4-mediated HIV-1 fusion can be inhibited by SDF-1. Because of its importance in the study of immunity and AIDS, large scale production of SDF-1 is desirable. In addition to recombinant technology, chemical synthesis provides means by which biologically active proteins can be produced not only in large quantity but also with a variety of designed modifications. In this study, we investigated the binding and function of an SDF-1α analogue, N33A, synthesized by a newly developed native chemical ligation approach. Radioiodinated N33A showed high affinity binding to human monocytes, T lymphocytes, as well as neutrophils, and competed equally well with native recombinant SDF-1α for binding sites on leukocytes. N33A also showed equally potent chemoattractant activity as native recombinant SDF-1α for human leukocytes. Further study with CXCR4/Fusin/LESTR transfected HEK 293 cells showed that N33A binds and induces directional migration of these cells *in vitro*. These results demonstrate that the chemically synthesized SDF-1α analogue, N33A, which can be produced rapidly in large quantity, possesses the same capacity as native SDF-1α to activate CXCR4-expressing cells and will provide a valuable agent for research on the host immune response and AIDS.

Stromal cell-derived factor (SDF)[1] 1 has been reported to be a primordial chemokine of the CXC subfamily and has multiple biological activities on a variety of cell types (1–4). SDF-1 was initially isolated as a T lymphocyte chemoattractant and was found to be active also on monocytes but not on neutrophils (1, 3). However, its activity on neutrophils was subsequently shown by using Ca^{2+} mobilization experiments (4). The SDF-1 gene is located on chromosome 10 (5), while the genes for other known CXC chemokines are located chromosome 4 (6, 7). SDF-1 is well conserved with only a single amino acid substitution between human and murirne molecules (1, 2). SDF-1α and SDF-1β were believed to be the result of differential splicing of a single gene, with SDF-1α missing four amino acids in the carboxyl terminus. The CXCR4/Fusin/LESTR has been identified as one of the functional receptors for SDF-1 (3, 4), a member of the seven transmembrane spanning receptor superfamily (8, 9). CXCR4 was initially cloned as an orphan receptor (10–12) and was later identified as a fusion co-factor for the entry of HIV-1 of the T lymphotropic strain (13). In addition to its activities on human leukocytes, SDF-1 has been shown to inhibit the fusion and replication of T lymphotropic HIV-1 in host cells bearing CD4 and CXCR4 (3, 4). Therefore, SDF-1 plays a pivotal role in host immune system and its defense against infection.

Rapid production of cytokines and chemokines is essential for structure-function studies and design of molecules with agonist or antagonist activities. The turbocharged peptide synthesis technology (14) and chemical ligation of peptide segments (15) permit the synthesis of chemokines in large quantity. This approach was therefore utilized to synthesize SDF-1α based on its great utility in studies of immune responses and infectious diseases. However, the thioacid peptide (SDF-1α(1–33)-αCOSH) could not be readily generated on the standard α-thiocarboxylate resin (16) due to the presence of an Asn^{33} residue at the ligation site in SDF-1α. Instead, alanine was coupled to the α-thiocarboxylate resin, then the mutant SDF-1α(1–33, N33A) was synthesized. In this study, we report that this SDF-1α analogue, N33A, displays full biological activity through the activation of the receptor CXCR4. This analogue, which can be rapidly produced in large quantity in the absence of contaminating peptides, will prove to be an important tool in the study of host immunity and AIDS.

MATERIALS AND METHODS

Chemokines—Human recombinant SDF-1α was purchased from PeproTech Inc. (Rocky Hill, NJ). Radioiodination of the SDF-1α and the analogue N33A was performed by a lactoperoxidase-labeling procedure. The radioactive ligands were further purified by reversed phase HPLC. The specific activity of the radioiodinated chemokines was 2200 Ci/mmol.

Chemical Synthesis of SDF-1α Analogue, N33A—Boc protected amino acids were obtained from the following sources: AnaSpec (San Jose, CA), Bachem (Philadelphia, PA), NovaBiochem (San Diego, CA), and Peptides International (Louisville, KY). Peptides were synthesized on a modified ABI430A instrument using *in situ* neutralization Boc chemistry protocols (17). C-terminal segments were prepared on –OCH2Pam resins (ABI, Foster City, CA). N-terminal segments were prepared on α-thiocarboxylate resin (16). Standard HF cleavage protocols were employed following N-terminal Boc removal and drying of the

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‖ To whom correspondence should be addressed: LMI, DBS, NCI-FCRDC, Bldg. 560, Rm. 31-19, Frederick, MD 21702. Tel.: 301-846-1347; Fax: 301-846-7042; E-mail: wangji@fcrfv1.ncifcrf.gov.

[1] The abbreviations used are: SDF, stromal cell-derived factor; HIV-1, human immunodeficiency virus type 1; HPLC, high pressure liquid chromatography; Boc, *t*-butoxycarbonyl; CI, chemotaxis index; rh, recombinant human.

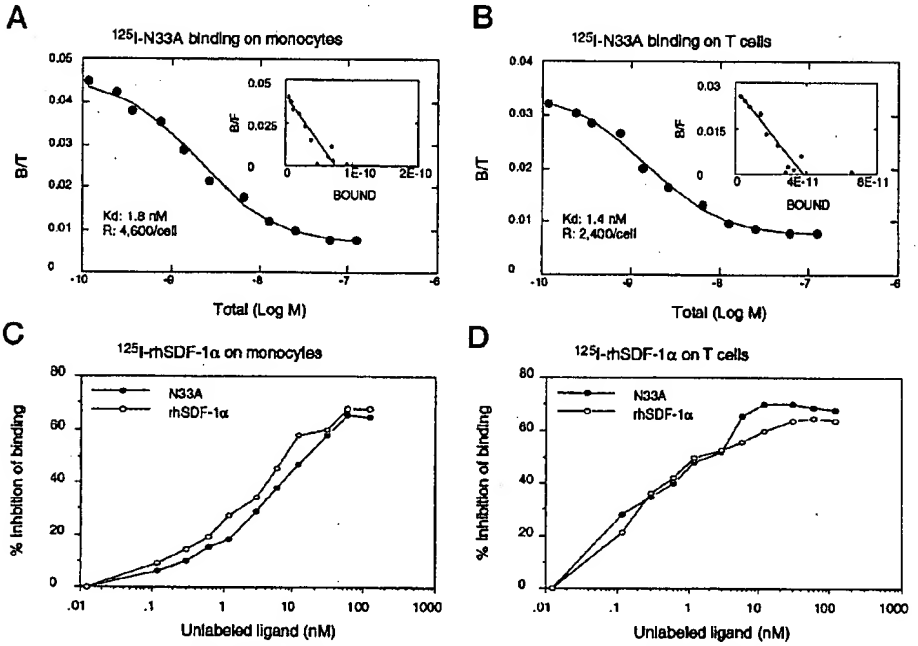


FIG. 1. **Binding of radiolabeled N33A to mononuclear cells and cross-competition with rhSDF-1α.** Cells were incubated for 40 min at room temperature with 0.12 nM radiolabeled N33A in the presence of increasing concentrations of unlabeled ligand. The cells were then harvested by centrifugation through a 10% sucrose phosphate-buffered saline cushion and measured for the radioactivity. The data were analyzed with LIGAND program. *Panel A*, binding of ^{125}I-N33A to monocytes; *panel B*, binding of ^{125}I-N33A to T cells; *panels C* and *D*, displacement of ^{125}I-SDF-1α binding to monocytes and T cells by unlableled ligands.

resin. HPLC purification was performed on Rainin HPLCs (Woburn, MA) using Vydac C4 or Dynamax C4 columns with gradient elution (A, H_2O, 0.1% trifluoroacetic acid; B, acetonitrile, 0.1% trifluoroacetic acid). Electrospray mass spectrometry was performed on a Sciex API1(PE-Sciex).

Ligation of the peptide segments was performed at 4 mM peptide concentration in 6 M guanidine, 0.1 M Tris, pH 7, in the presence of 33 mM thiophenol (Fluka, Switzerland) at room temperature. Ligation was monitored by HPLC and was typically complete within 24 h. Ligation was followed by HPLC purification and lyophilization. After purification, the full-length peptide was reduced at 1 mg/ml in 8 M urea (Fluka), 0.1 M Tris (Fluka), 5.37 mM EDTA ((Fluka), pH 8.6, in the presence of 100 mM 2-mercaptoethanol (Fluka). Reduction occurs under a nitrogen atmosphere at 40 °C for 1 h. After complete reduction, the mixture was reconstituted with the same buffer at 0.2 mg/ml with 18.7 mM oxidized glutathione (Sigma). The solution was dispensed into a Spectrum Spectra/Por*7 dialysis membrane (Houston, TX) (M_r cut-off, 3500) and the bag was placed in 1 liter of initial dialysis buffer of 8 M urea, 0.1 M Tris, 1 mM EDTA, 3 mM 2-mercaptoethanol, 1.3 mM oxidized glutathione, pH 8.6. Over a period of 2 days, 4 liters of 2 M urea, 0.1 M Tris, pH 8.6, was pumped into the vessel containing the dialysis bag (18). After lyophilization, the full-length peptide was oxidized at 1 mg/ml in 2 M guanidine HCl (Fluka), 0.1 M Tris (Fluka), pH 8.6, at room temperature in the presence of air. Folding was complete after stirring overnight and was monitored by HPLC and mass spectrometry.

Cells and Chemotaxis Assays—Human peripheral blood leukocytes were isolated from normal donors (National Institutes of Health Clinical Center Transfusion Department, Bethesda, MD) according to the established protocols in this laboratory for purification of monocytes (purity was >90%) (19), T lymphocytes (purity, >95% $CD3^+$) (20) and neutrophils (purity, >98%) (19). The chemokine receptor, CXCR4/fusin/LESTR cDNA was isolated in this laboratory and was stably transfected into human kidney embryonic epithelial 293 cells (CXCR4/293 cells) as described previously (21).

Leukocyte migration was evaluated using a 48-well microchamber (Neuroprobe, Cabin John, MD) technique as described previously (19, 20, 22). The migration of CXCR4/293 cells was also assessed by the 48-well microchamber technique with the polycarbonate filters (10-μm pore size) (21) precoated with collagen type I (Collaborative Biomedical Products, Bedford, MA). The results are expressed as the the chemotaxis index (CI) representing the fold increase in the cell migration induced by stimuli *versus* control medium. All experiments were performed at least two times, and results from one experiment are shown. The statistical significance of the difference between migration in response to stimuli and control was assessed by Student's *t* test.

Binding Assays with Radiolabeled SDF-1α and N33A—Binding assays were performed using a single concentration of ^{125}I-labeled chemokines in the presence of increasing concentrations of unlabeled ligands (19, 21). The binding data were analyzed with a Macintosh computer program LIGAND (P. Munson, Division of Computer Research and Technology, NIH, Bethesda, MD). In Scatchard plots, the binding data were analyzed with both "one-site" and "two-site" models, and only the one-site model better fit the curves obtained with either native leukocytes or CXCR4/293 cells. The rate of competition for binding by unlabeled ligands was calculated with the following formula:

$$\% \text{ inhibition} = 1 - (\text{binding in the presence of unlabeled chemokine/binding in the presence of medium alone}) \times 100 \quad \text{(Eq. 1)}$$

RESULTS AND DISCUSSION

We first examined whether the SDF-1α analogue N33A was able to bind and activate human peripheral blood monocytes and T cells. Fig. 1 shows that ^{125}I-N33A specifically bound to human peripheral blood monocytes (Fig. 1*A*) and T lymphocytes (Fig. 1*B*) with high affinity (1.8 and 1.4 nM for monocytes and T cells, respectively). This level of binding to monocytes and T cells by ^{125}I-N33A was comparable to that of ^{125}I-rhSDF-1α as examined in parallel experiments (data not shown). Unlabeled N33A displaced ^{125}I-rhSDF-1α binding to both monocytes and T lymphocytes (Fig. 1, *C* and *D*), and likewise, the binding of ^{125}I-N33A to these cell types was also displaced by both unlabeled N33A and rhSDF-1α (not shown). Consistent and considerable migration of monocytes and T cells was induced by N33A (Fig. 2). The potency and efficacy of N33A in the induction of mononuclear cell migration was comparable with rhSDF-1α, suggesting that chemically synthesized N33A retains the tertiary structure and functions as well as rhSDF-1α.

The effect of SDF-1 on neutrophils is controversial (1, 3, 4). While some investigators failed to detect chemotactic activity of SDF-1 for neutrophils (1), others were able to induce significant Ca^{2+} mobilization in neutrophils at physiologically relevant concentrations of SDF-1 (4). In an effort to clarify the activity of SDF-1 on neutrophils, we tested the binding and function of

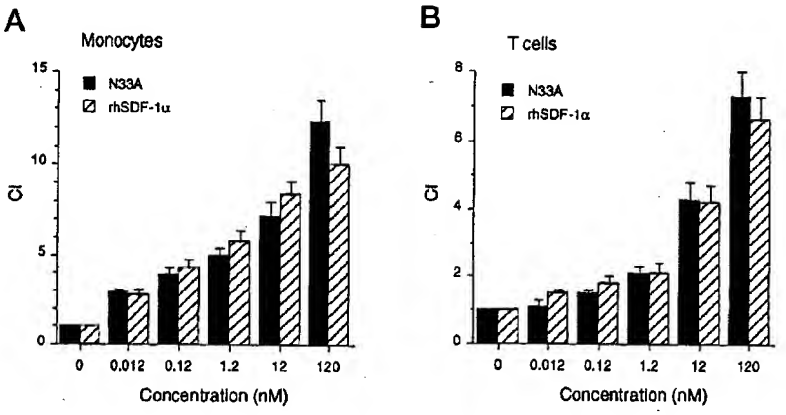


FIG. 2. **Chemotactic response of monocytes and T cells to N33A and rhSDF-1α.** Different concentrations of N33A or rhSDF-1α were placed in the lower wells of the microchemotaxis chamber. The cells were placed in the upper wells, which were separated from the lower wells by polycarbonate filters. After incubation the filters were removed, stained, and the cells that migrated across the filters were counted. The results are expressed as the CI representing the fold increase of cell migration in response to chemokines *versus* medium control. CI ≥ 2 are statistically significant in comparison with the spontaneous migration (in response to control medium alone, CI = 1).

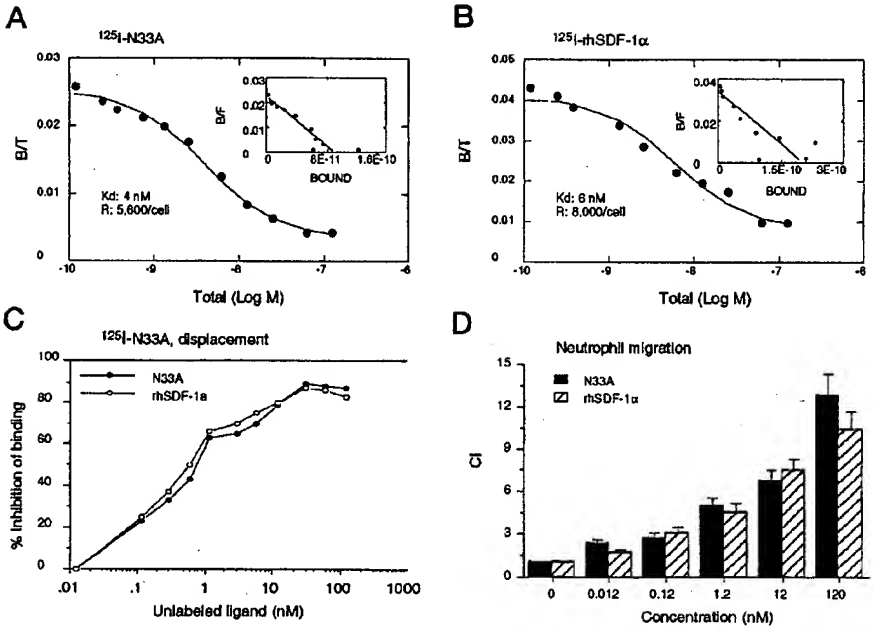


FIG. 3. **Effect of N33A and rhSDF-1α on human peripheral blood neutrophils.** The binding of radiolabeled ligands to neutrophils and the chemotactic responses were examined as described for monocytes and T cells. *Panel A*, binding of N33A to neutrophils; *panel B*, binding of rhSDF-1α to neutrophils; *panel C*, displacement of ^{125}I-N33A binding to neutrophils by unlabeled ligands; *panel D*, neutrophil migration in response to N33A and rhSDF-1α. CI ≥ 2 are statistically significant in comparison with the spontaneous migration.

N33A on neutrophils in comparison rhSDF-1α. As shown in Fig. 3, human peripheral blood neutrophils expressed a substantial number of specific binding sites for both N33A (Fig. 3*A*) and native SDF-1α (Fig. 3*B*). The binding is of high affinity with estimated K_d values of about 5 nM. N33A and rhSDF-1α competed equally well for each other's binding as shown by the displacement curve (Fig. 3*C* and data not shown). Neutrophils also migrated in response to both N33A and rhSDF-1α, indicating that neutrophils are indeed among the target cell types for SDF-1.

To further confirm that N33A utilizes CXCR4 as its functional receptor, HEK293 cells stably expressing CXCR4 (CXCR4/293 cells) were employed. Wild type HEK 293 cells exhibited a low level of specific binding (about 200 binding sites/cell) for both N33A and native SDF-1α. Both N33A and native SDF-1 also induced a weak but significant directional migration of wild type HEK293 cells (CI = 2.1 ± 0.2, at 120 nM ligand concentration). This is in agreement with the notion that a great variety of nonhematopoietic cells express CXCR4 mRNA. However, HEK293 cells overexpressing CXCR4 expressed markedly increased number of specific binding sites for radiolabeled N33A and rhSDF-1α (Fig. 4, *A* and *B*). N33A and rhSDF-1α mutually competed for binding to CXCR4/293 cells (Fig. 4, *C* and *D*). Both N33A and rhSDF-1α were able to induce a remarkable directional migration of CXCR4/293 cells (Fig. 5) with similar potency and efficacy. HEK293 cells transfected with known chemokine receptors, including CXCR1, CXCR2, and CCR1–5, did not show any increased binding or activation by N33A or rhSDF-1α over background levels, whereas these cells specifically bound and migrated in response to their ligands (data not shown). These results demonstrate that N33A, like rhSDF-1α, uses CXCR4 as a functional receptor.

Chemokines are important mediators that participate in a variety of pathophysiological conditions including inflammation, infection, tissue injury and repair, immune responses, as well as malignancy (6, 7). Recently members of chemokine receptor family have been identified to be fusion co-factors for HIV-1 entry into host cells. Some chemokines were able to inhibit HIV-1 entry through competitive occupancy of the relevant receptors such as CCR5 or CXCR4/Fusin/LESTR which mediate the fusion of either monocytotropic or T lymphocytotropic viruses, respectively, with $CD4^+$ host cells (2, 3, 13, 23–27). SDF-1 has recently been identified as the ligand for CXCR4 and was able to inhibit the cell fusion mediated by the envelope protein of the T lymphotropic HIV-1 strain (23–28).

SDF-1 is a member of the CXC chemokine subfamily and was first identified as a molecule possessing pre-B cell-stimulatory activity (29, 30). It was later described as a chemoattractant for resting T lymphocytes and monocytes (1). As a CXC chemokine, SDF-1 has several unique features in comparison to other

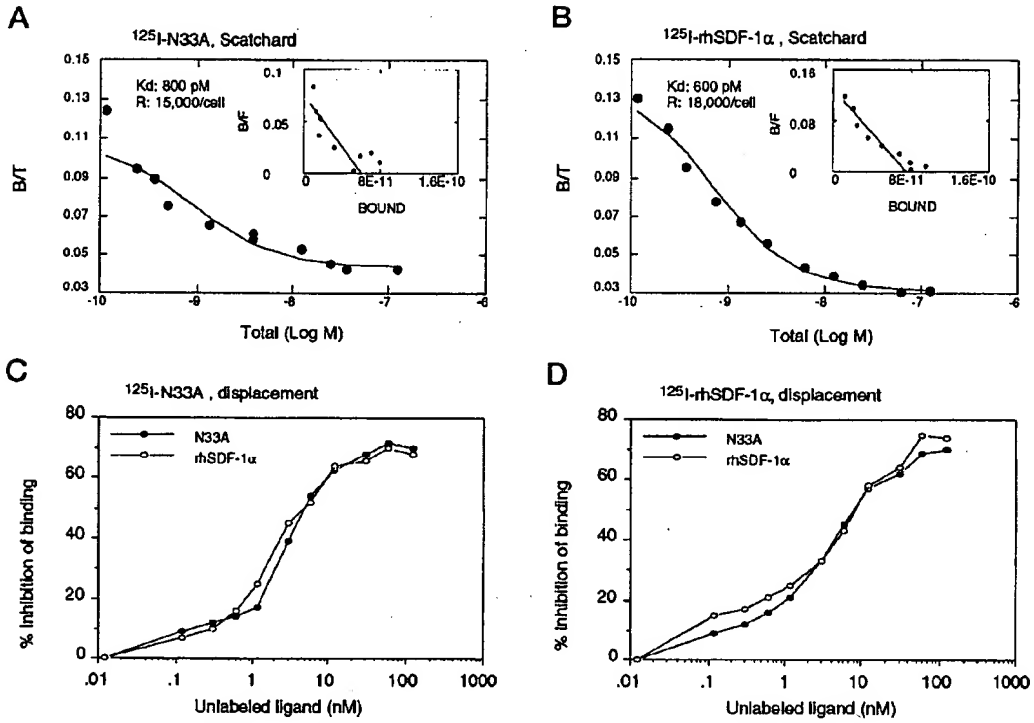


FIG. 4. **Binding of N33A and rhSDF-1α to CXCR4/293 cells.** CXCR4/293 cells were gently detached from plastic flasks with trypsin/EDTA, washed, and measured for their ability to bind N33A and rhSDF-1α. *Panel A*, binding of N33A to CXCR4/293 cells; *panel B*, binding of rhSDF-1α to CXCR4/293 cells; *panel C*, displacement of N33A binding to CXCR4/293 cells by unlabeled ligands; *panel D*, displacement of rhSDF-1α binding to CXCR4/293 cells by unlabeled N33A and rhSDF-1α.

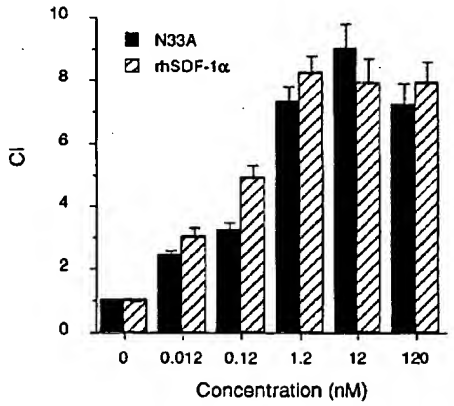


FIG. 5. **Chemotactic response of CXCR4/293 cells induced by N33A and rhSDF-1α.** CXCR4/293 cell migration was measured by 300 min incubation at 37 °C in microchemotaxis chambers as described previously (21). CI ≥ 2 are statistically significant in comparison with the spontaneous migration.

members of the same family. SDF-1 mRNA is expressed constitutively in virtually every tissue including heart, liver, lung, brain, muscle, spleen, and kidney (28–30). Expression of SDF-1 gene is not affected by proinflammatory stimuli (31), in contrast to most other chemokines which are mainly expressed in response to proinflammatory cytokines and are believed to regulate the recruitment and activation of mature leukocytes at inflammatory foci (6, 7). In the absence of inflammation, blood monocytes constantly replace mononuclear phagocytes in the tissue, sustaining a stable level by extravasation from blood stream and undergoing differentiation into macrophages. SDF-1 displays a tissue distribution that is considered appropriate for function in lymphocyte recirculation, in basal recruitment of monocytes and in normal replenishment and turnover of tissue mononuclear phagocytes (1, 30–32). However, neutrophils do not normally infiltrate organs or tissues even though these cells express binding sites for and migrate in response to SDF-1α *in vitro*, as further demonstrated in the current study. Thus, the role for SDF-1 as a primordial chemokine regulating primarily the tissue distribution of leukocytes needs further investigation.

SDF-1 also has several essential functions in development (2). Mice lacking SDF-1 due to homologous recombination died perinatally and although the numbers of B cell progenitors in mutant embryos were severely reduced in fetal liver and bone marrow, myeloid progenitors were reduced only in the bone marrow not in the fetal liver, indicating that SDF-1 is responsible for B cell lymphopoiesis and bone marrow myelopoiesis. In addition, mice deprived of SDF-1 gene had a cardiac ventricular septal defect (2).

The SDF-1α analogue N33A was chemically synthesized and was shown in this study to be equally as potent as native SDF-1α in binding and activating human leukocytes as well as CXCR4-transfected HEK293 cells. N33A has also been shown to induce human B lymphocyte migration and to effectively inhibit infection of PM-1 cells by HIV-1(LAV).[2] The preparation of analogues is greatly facilitated by molecular chemical synthesis, in which proteins can be produced either singly or by combinational methods. The analogues could include a full range of genetically encoded amino acids as well as unnatural backbone structures and unclonable residues such as D-amino acids, fluorescent or nuclear magnetic resonance-sensitive nuclei. The activities of analogues can also be tuned by fast cycles of synthesis-design-assay-resynthesis (33). The validity of this approach is demonstrated by the fact that it yields functional molecules such as the SDF-1α analogue N33A, a biological

[2] M. A. Siani, D. A. Thompson, L. E. Canne, G. M. Figliozzi, S. B. H. Kent, R. Simon, J. Cyster, P. E. Kennedy, E. D. Smith, and E. A. Berger, unpublished observation.

contamination-free agonist of SDF-1 that can be utilized in studies of the immune system and host defense against AIDS.

Acknowledgment—We thank Kathleen Bengali and Nancy Dunlop for technical support, and Cheryl Fogle and Teresa Covell for secretarial assistance. The critical review of this manuscript by Dr. Joost J. Oppenheim is greatly appreciated.

REFERENCES

1. Bleul, C. C., Fuhlbrigge, R. C., Casasnovas. J. M., Aiuti, A., and Springer, T. A. (1996) *J. Exp. Med.* **184,** 1101–1109
2. Nagasawa, T., Hirota, S., Tachibana, K., Takakura, N., Nishikawa, S., Kitamura, Y., Yoshida, N., Kikutani, H., and Kishimoto, T. (1996) *Nature* **382,** 635–638
3. Bleul, C. C., Fazzan, M., Choe, H., Parolin, C., Clark-Lewis, I., Sodroski, J., and Springer, T. A. (1996) *Nature* **382,** 829–833
4. Oberlin, E., Amara, A., Bachelerie, F., Bessia, C., Virelizier, J-L., Arenzana-Seisdedos, F., Schwartz, O., Heard, J.-M., Clark-Lewis, I., Legler, D. F., Loetscher, M., Baggiolini, M., and Moser, B. (1996) *Nature* **382,** 833–835
5. Shirozu, M., Nakano, T., Inazawa, J., Tashiro, K., Tada, H., Shinohara, T., and Nonjo, T. (1995) *Genomics* **28,** 495–500
6. Oppenheim, J. J., Wang, J. M., Chertov, O., Taub, D. D. and Ben-Baruch, A. (1996) in Tilney, L. N., Strom, T. B., and Paul, L. C. (eds) *Transplantation Biology: Cellular and Molecular Aspects*, Lippincott-Raven, Philadelphia, pp. 187–200
7. Baggiolini, M., Dewald, B., and Moser, B. (1994) *Adv. Immunol.* **55,** 97–179
8. Kelvin D. J., Michiel, D. F., Johnston, J. A., Lloyd, A. R., Sprenger, H., Oppenheim, J. J., and Wang, J. M. (1993) *J. Leukocyte Biol.* **54,** 604–612
9. Murphy, P. M. (1996) *Cytokine Growth Factor Rev.* **7,** 47–64
10. Loetscher, M., Geiser T., O'Reilly, T., Zwahlen, R., Baggiolini, M., and Moser, B. (1994) *J. Biol. Chem.* **269,** 232–237
11. Federsppiel, B., Melhado, I. G., Duncan, A. M. V., Delaney, A., Schappert, K., Clark-Lewis, I., and Jirik, F. R. (1993) *Genomics* **16,** 707–712
12. Nomura, H., Nielsen, B. W., and Matsushima, K. (1993) *Int. Immunol.* **5,** 1239–1249
13. Feng, Y., Broder, C. C., Kennedy, P. E., Berger, E. A. (1996) *Science* **272,** 872–877
14. Schnolzer, M., Alewood, P., Jones, A, Alwood, D., and Kent, S. B. H. (1992) *Int. J. Pept. Protein Res.* **40,** 180–193
15. Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent S. B. H. (1994) *Science* **266,** 776–779
16. Canne, L. E., Walker, S. M., and Kent, S. B. H. (1995) *Tetrahedron Lett.* **36,** 1217–1220
17. Canne, L. E., Ferre-D'Amare, A. R., Burley, S. K., and Kent S. B. H. J. (1995) *J. Am. Chem. Soc.* **117,** 2998–3007
18. Maeda, Y., Ueda, T., and Imoto, T. (1996) *Protein Eng.* **9,** 95–100
19. Xu, L. L., MacVicar, D. W., Ben-Baruch, A., Kuhns, D. B., Johnston, J., Oppenheim, J. J., and Wang, J. M. (1995) *Eur. J. Immunol.* **25,** 2612–2617
20. Xu, L., Kelvin, D., Ye, G. Q., Taub, D. D., Ben-Baruch, A., Oppenheim, J. J., and Wang, J. M. (1995) *J. Leukocyte Biol.* **57,** 335–342
21. Ben-Baruch, A., Xu, L., Young, P. R., Bengali, K., Oppenheim, J. J., and Wang, J. M. (1995) *J. Biol. Chem.* **270,** 22123–22128
22. Falk, W. R., Goodwin, R. H., Jr., Leonard, E. J. (1980) *J. Immunol. Methods.* **33,** 239–247
23. Deng, H. K., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Marzio, P. D., Marmon, S., Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C., Schall, T. J., Littman, D. R., and Landau, N. R. (1996) *Nature* **381,** 661–666
24. Dragic, T., Litwin, V., Allaway, G. P., Martin, S., Huang, Y., Nagashima, K. A., Cayaman, C., Maddon, P. J., Koup, R. A., Moore, J. P., and Paxon, W. A. (1996) *Nature* **381,** 667–673
25. Alkhatib, G., Combadiere, C., Broder, C. C., Feng, Y., Kennedy, P. E., Murphy, P. M., and Berger, E. A. (1996) *Science* **272,** 1955–1958
26. Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P, D., Wu, L., Mackay, C. R., LaRosa, G., Newman, W., Gerard, N., Gerard, C., and Sodroski, J. (1996) *Cell* **85,** 1135–1148
27. Doranz, B. J., Rucker, J., Yi, Y., Smyth, R. J., Samson, M., Peiper, S. C., Parmentier, M., Collman, R. G., and Doms, R. W. (1996) *Cell* **85,** 1149–1158
28. Cocchi, F., DeVico, A. L., Garzino-Demo, A., Arya, S. K., Gallo, R. C., and Lusso, P. (1995) *Science* **270,** 1811–1815
29. Nagasawa, T. Kikutani, H., and Kishimoto, T. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 2305–2309
30. Tashiro, K. Tada, H., Heilker, R., Shirozu, M., Nakano, T., and Honjo, T. (1993) *Science* **261,** 600–603
31. Jiang, W., Zhou, P., Kahn, S. M., Tomita, N., Johnson, M. D., and Weinstein, I. B. (1994) *Exp. Cell Res.* **215,** 284–293
32. Godiska, R., Chantry, D., Dietsch, G. N., and Gray, P. W. (1995) *J. Neuroimmunol.* **58,** 167–176
33. Robson, B. (1996) *Nat. Biotechnol.* **14,** 892–893

Chemical protein synthesis

Jill Wilken and Stephen BH Kent

Chemical protein synthesis is a field in transition. Previously, the synthetic accomplishment itself was the major focus of work in this field. Increasingly, chemical synthesis is now being applied to understanding how biological function originates in the structure of the protein molecule. A novel approach – 'chemical ligation', which is the chemoselective reaction of unprotected peptide segments in water at pH7 – has made the total synthesis of proteins a robust and practical route to the study of structure-function relationships. For certain protein families, chemical protein synthesis is the most effective way to obtain functional proteins direct from genome sequence data.

Address
Gryphon Sciences, South San Francisco, California, USA

Current Opinion in Biotechnology 1998, 9:412–426

http://biomednet.com/elecref/0958166900900412

 ISSN 0958-1669

Abbreviations

Aba	L-α-NH_2-*n*-butyric acid
ESMS	electospray ionization mass spectrometry
rec	recombinant
SPPS	stepwise solid phase peptide synthesis

Introduction

Proteins are the functional gene products, responsible for essentially all the activities of the biological world. For this reason, an understanding of protein structure and function is an important objective of modern biomedical research. In recent years, proteins have been most commonly studied by over-expression in microorganisms using recombinant (rec) DNA-based technology [1]. Techniques such as site-directed mutagenesis [2] have allowed the systematic variation of protein structure and have vastly increased our understanding of the protein world.

The preparation of protein molecules by chemical synthesis has long been an objective of researchers interested in the molecular origins of protein function [3]. The goal of total chemical synthesis of protein molecules provided an incentive to classical organic chemists over a period of decades and culminated in the synthesis of insulin and a series of disulfide-mispaired (and hence deliberately misfolded) insulin analogues [4]. More recently, Sakakibara [5] has persevered with the classic solution chemistry approach to the total synthesis of proteins, by means of convergent chemical condensation of fully protected peptide segments in organic solvents. This technique has scored a number of successes, such as the total synthesis of midkine [6] and related molecules, and has been used for study of structure-activity relationships [7].

Merrifield's stated goal in developing the stepwise solid phase peptide synthesis (SPPS) method was the total chemical synthesis of proteins [8]. In the SPPS technique, protected amino acids in organic solvents are added one at a time to a resin-bound peptide chain, resulting in the assembly of the target peptide in fully-protected, resin-bound form. The product polypeptide is released by deprotection and cleavage from the resin support. Perhaps the ultimate accomplishment of the stepwise SPPS method was Schneider's total synthesis of crystalline HIV-1 protease [9]. Subsequently, Wlodawer *et al.* [10] used this chemically synthesized protein to obtain the first correct crystal structure of this enzyme. Interestingly, Miller *et al.* [11,12] obtained the first co-crystal structures of the HIV-1 protease with substrate-derived inhibitors bound to the enzyme molecule, using protein prepared by total chemical synthesis. The resulting structural data were made freely available to the research community and formed the basis of the world-wide structure-based design programs that led to the development of the highly successful 'protease inhibitor' class of AIDS therapeutics [13].

Wells and co-workers [14] have used enzymatic ligation of unprotected peptide segments for the total chemical synthesis of proteins, developing enzymes specifically engineered for this purpose. Thus, in a classic study, RNase A and analogues were prepared using 'subtiligase' (an engineered form of subtilisin). In the current review period, Wong and co-workers [15•] have used enzymatic peptide re-ligation and enzymatic glycosylation for the production of ribonuclease B glycoforms.

The search for more effective routes to the understanding of protein function has gained added urgency, because of the success of the expressed sequence tag genome projects. Huge numbers (>100,000) of 'proteins' have been discovered in the past two or three years, but are at present frequently known only as open reading frames in a computer sequence database [16]. This has led to a research bottleneck that was unanticipated as recently as three years ago; how can biomedical researchers obtain proteins, the 'functional gene products', from database sequences in a timely and cost-effective manner?

Fortunately, a new technology has emerged that suggests that this unanticipated problem has an equally unlikely answer: total chemical protein synthesis! This review selectively focuses on this most promising of new protein technologies, the synthesis of functional proteins directly from genome sequence data by 'chemical ligation', that is, the chemoselective reaction of unprotected peptide segments in water at pH7.

Functional proteins directly from genome sequence data

As discussed above, the HIV-1 protease molecule was one of the first examples of the use of chemical protein synthesis to gain direct access to the functional protein, when all that was known was the putative polypeptide sequence encoded in genome data [9].

More recently, Clark-Lewis *et al.* [17] have very effectively used total chemical protein synthesis by the stepwise SPPS method (Figure 1) to establish much of the basic knowledge in the chemokine protein family [17]. Again, the approach has been to prepare the proteins by total chemical synthesis based on DNA sequence data. In addition to providing direct access to functional wild-type proteins, Clark-Lewis *et al.* [17] have also made unique analogues to study important aspects of chemokine biology. Notably, in order to study the functional role of protein dimer formation they constructed an 'obligate monomer' form of IL-8 [18••]. This chemical protein analogue had a methylated backbone amide bond that prevented the formation of dimeric protein in solution, and yet retained full biological activity [19].

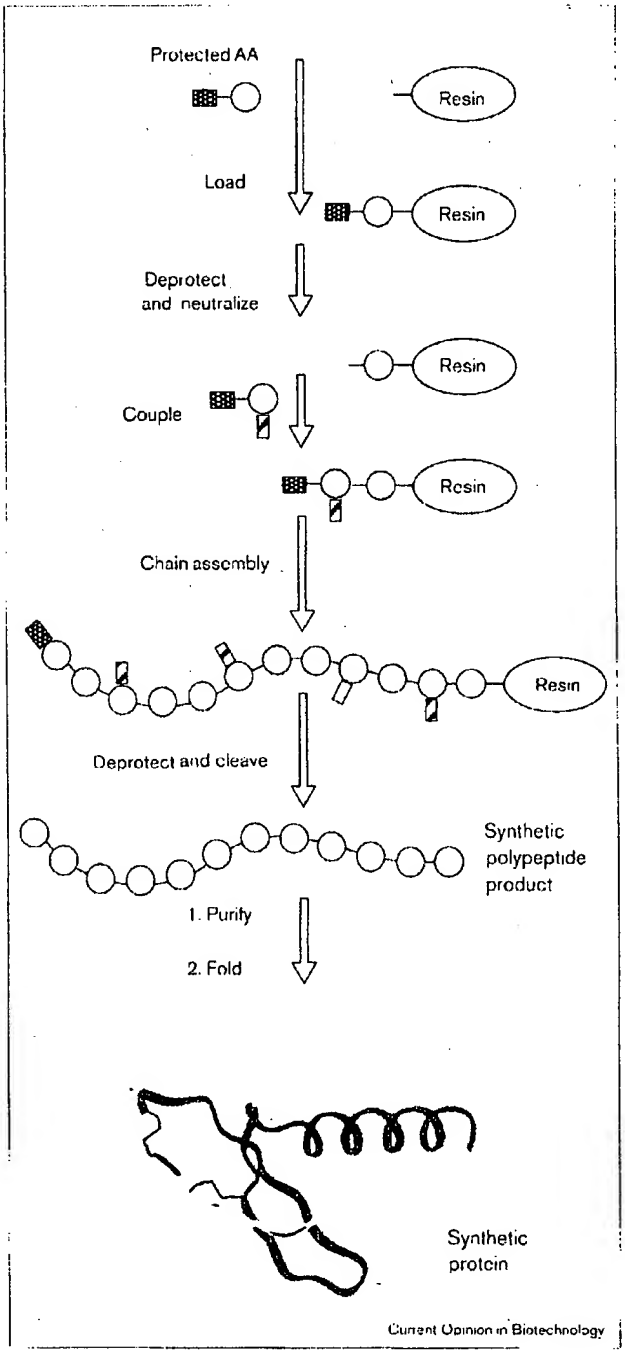
Chemical ligation

The most important recent development in the total synthesis of proteins is the chemical ligation method [20]. The underlying concept of this novel approach is shown in Figure 2. Chemical ligation uses the principle of chemoselective reaction — that is, the reaction of two large (35–60 amino acid residue) unprotected peptide segments in water at neutral pH, to yield product polypeptides up to about 120 amino acids in length. An unnatural moiety replaces the native peptide bond at the site of ligation. A chaotropic agent, such as 6 M guanidine.HCl, is used to increase the concentration of the reacting peptide segments to 2–5 mM, insuring rapid and complete reaction. Quantitative ligation is typically obtained in a matter of hours, and the polypeptide product is simply isolated in one step by preparative reverse phase HPLC [21].

Use of unprotected peptide segments has a number of advantages. Synthetic peptides up to ~60 amino acids can be readily prepared by highly optimized stepwise SPPS [22]. Unprotected peptides can be purified to a high degree of homogeneity by standard techniques, such as reverse phase HPLC or ion exchange chromatography. The entire process, including both chain assembly and purification, takes less than 24 hours. Furthermore, unprotected peptides can be characterized very effectively by techniques such as electrospray mass spectrometry which rely on ionizable (i.e. unprotected) functional groups for their effectiveness. Finally, unprotected peptides frequently have good solubility in aqueous solution; if necessary, their solubility can be dramatically enhanced by the use of a chaotropic agent such as 6M guanidine.HCl [20].

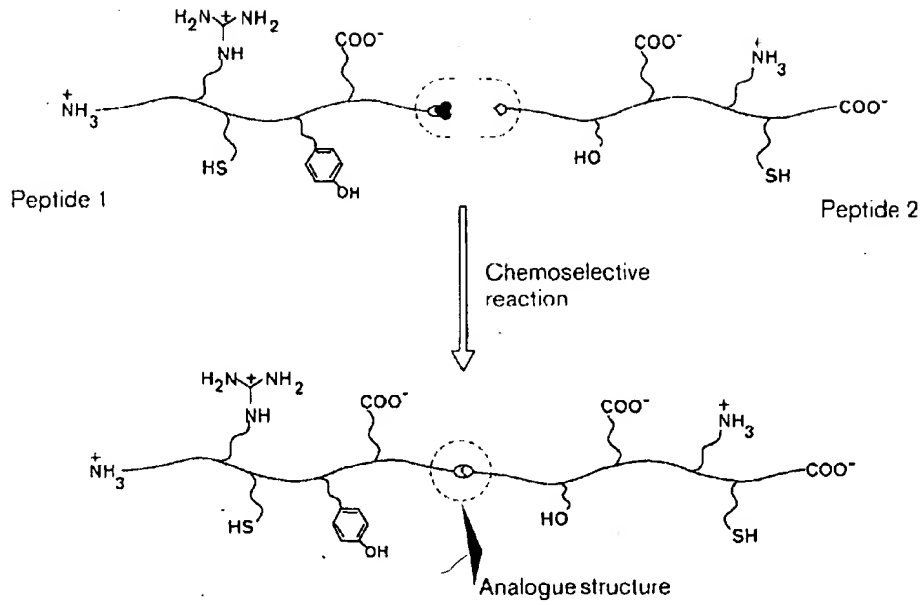
The chemical ligation approach appears to be generally applicable. It has been used to make a wide range of protein molecules, ranging up to 22 kDa in size [23] (Table 1). Proteins prepared by total chemical synthesis in this manner have the expected biological activities, and are correctly folded. In several examples, techniques such as NMR [24] and X-ray crystallography [25••] have been used to

Figure 1



Chemical protein synthesis by means of stepwise SPPS. Because of the accumulation of resin-bound by products, the size of polypeptide that can be synthesized in this way is restricted. Stepwise SPPS is limited to the synthesis of small proteins. AA, amino acid residue.

Figure 2



Principles of chemical ligation. Two completely unprotected peptide segments are covalently joined by means of the chemoselective reaction of unique, mutually reactive functional groups. In the original version of this approach [20], an unnatural (i.e. non-peptide bond) moiety is formed at the ligation site.

elucidate the three dimensional structure of proteins prepared by total chemical synthesis, using the chemical ligation method in its original form.

Another interesting example of the use of the chemical ligation method in its original form to elucidate novel aspects of the molecular basis of protein function, is the total synthesis of a covalent heterodimer made up of the b/HLH/Z regions of the transcription factors cMyc and Max [26]. Canne *et al.* [26] used convergent chemical ligation of four large peptide segments to make a synthetic protein construct that, although a polypeptide chain with linear topology, had two amino-terminals and no carboxy-terminal! This fused bi-directional polypeptide folded efficiently to give a protein of 20.6 kDa with the expected DNA-binding activities [26].

Table 1

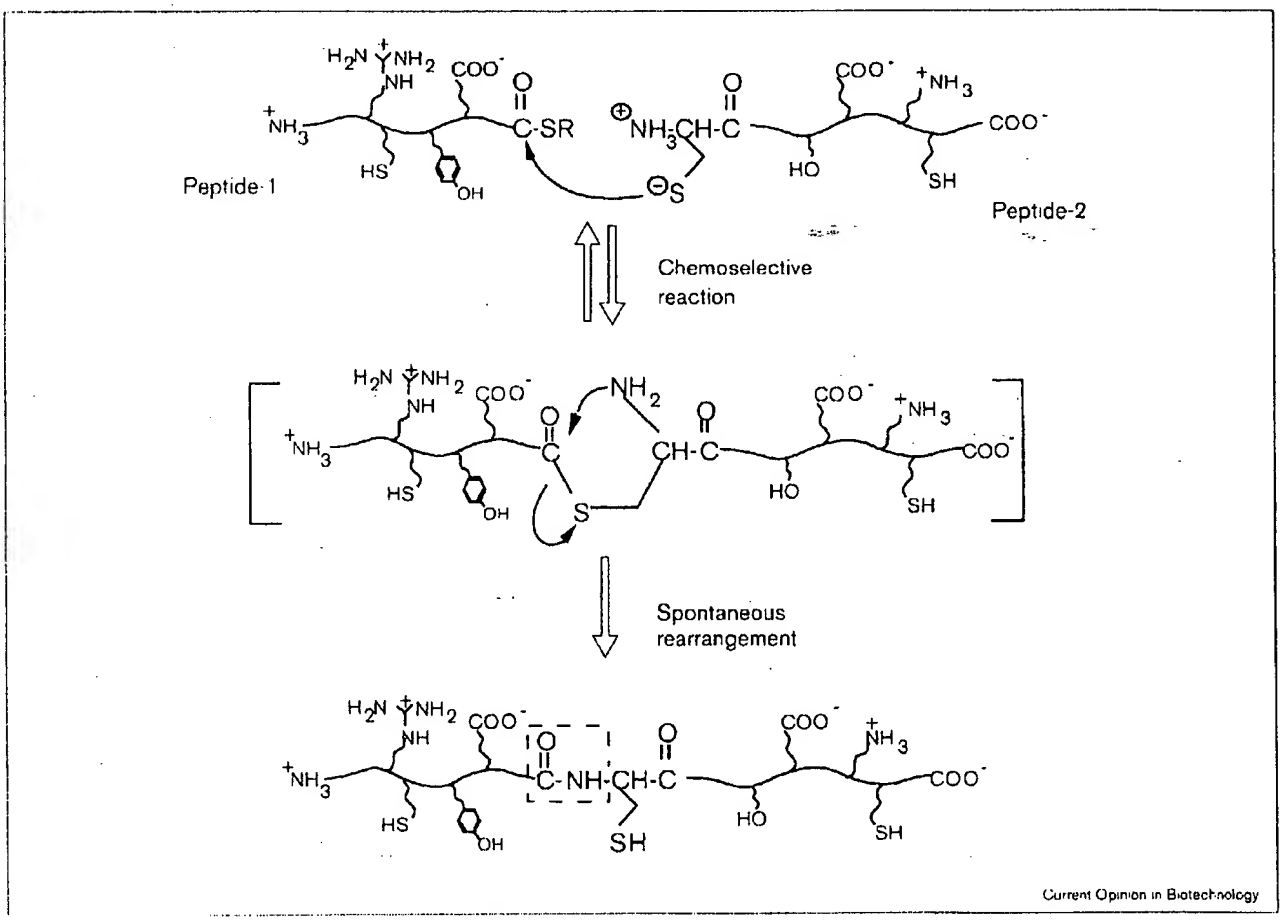
Proteins prepared by chemical ligation

Protein	Size No. amino acids	References
b/HLH transcription factors	180	[26]
HIV-1 protease	2 × 99	[20]
D-HIV-1 protease	2 × 99	[21]
Flap engineered HIV-1 protease	2 × 99	[54]
HIV-1 protease (tethered dimer)	202	[23]
Intergrin 2b3a neoprotein	126	[89]
Fibronectin 10F3 domain	94	[24]
Cytochrome *b* TASP	122	[69•]
Artificial proteins	195	[67]

Native chemical ligation

In 1994, Dawson *et al.* [27] introduced a remarkable refinement of the chemical ligation method in which two unprotected peptide segments were reacted to directly yield a polypeptide product with a native peptide bond at the site of ligation. 'Native chemical ligation' as this technique is called, has proven to be a robust and practical method for the total synthesis of proteins. The principles of native chemical ligation are shown in Figure 3. The method involves thioester-mediated chemoselective reaction of two unprotected peptide segments in aqueous solution at neutral pH, and results in ligation at a cysteine residue *via* a native amide bond. A typical example is the total synthesis of the enzyme human secretory phospholipase A2 (124 amino acids, 14 cysteine residues in seven disulfides) [28•]. The native chemical ligation method has been the subject of mechanistic studies [29••], and has been reviewed [30]. Ever increasing numbers of proteins have been made using this technique (Table 2). Because of its speed and efficiency for small, cysteine-rich proteins, native chemical ligation has become the method of choice for the rapid preparation of wild-type (i.e. native) proteins directly from genome sequence data [31•].

Figure 3



Native chemical ligation. In a refined version of the chemical ligation approach, two unprotected peptides are reacted in aqueous solution at pH 7 to yield a polypeptide product with a native peptide bond at the ligation site [27]. This is accomplished by thiol exchange between the carboxy-terminal thioester of one peptide and the thiol side chain of the cysteine residue at the amino-terminal of the other peptide. Rapid, irreversible intramolecular nucleophilic attack yields the native peptide bond.

The protein folding 'problem'

Chemical synthesis of large polypeptide chains is a necessary step in the total synthesis of proteins, but is not in itself sufficient: the polypeptide product must be accurately folded to form the unique three-dimensional structure of the protein molecule. It is this three-dimensional structure that determines the activity of the protein. Several decades ago, Anfinsen [32] hypothesized that the linear sequence of amino acids in a polypeptide chain was in itself sufficient to determine the precise three-dimensional structure of the protein molecule. Abundant evidence from *in vitro* folding studies of isolated proteins has supported this hypothesis. Indeed, because it is the linear amino acid sequence alone that is encoded by the DNA sequence of the gene, it must of necessity determine the functional form of the protein.

More recently, an elaborate cellular machinery associated with protein folding *in vivo* has been elucidated [33]. Although none of the research into this 'chaperone' apparatus has been interpreted as contradicting Anfinsen's fundamental assertion, it has nonetheless come to be widely believed that special cellular conditions are required for accurate protein folding, especially of larger multi-domain proteins. Such skepticism has been reinforced by the frequent observation that proteins that are recombinantly expressed at high levels in microbial cells form inclusion bodies [34] (i.e. intracellular precipitates); proteins recovered from these inclusion bodies are often difficult to fold efficiently to give active protein molecules [35]. This mind-set has caused considerable skepticism about the possibility of accurate *in vitro* folding of chemically synthesized large polypeptides to form native, functional protein molecules.

Concerns about the correct folding of synthetic polypeptide chains can only be effectively answered by experiment. So far, in the chemical synthesis of more than 200

Table 2

Proteins prepared by native chemical ligation.

Class/family	Size No. amino acids	No. of Cys [disulfides]	References
Secretory			
Chemokines	68–90	4 [2]	[31•,48]
Ser PR inhibitors	58–70	6 [None]	[37,41,55•]
AGRPs	46–112	10 [5]	[D Thompson]
Anaphlyatoxins	72	6 [3]	[R Simon]
EGFs/TGF-α	50	6 [3]	[M Siani]
Receptor/membrane			
β2 Microglobulin	99	2 [1]	[J Wilken]
Intracellular			
SH2 domains	95	1* [None]	[T Muir]
SH3 domains	60	1*	
Transcription factor	72	6 [None]	[J Wilken]
Redox			
Desulforedoxin	2 × 36	4 [None]	[D Low]
Rubredoxin	53	4 [None]	
Cytochrome b5	82	1* [None]	[C Hunter]
Enzymes			
Retroviral proteases	2 × (99–116)	2 [None]	[S Walker]
Human secPLA2 (G2)	124	14 [7]	[28•]
Macrophage migration inhib. factor	3 × 115	2 [None]	[L Canne]
Barnase	110	1* [None]	[29••]

*No. native cysteine residues; a single cysteine introduced to form a ligation site.

proteins from numerous families and falling into many different classes (Tables 1 and 2), efficient folding of synthetic polypeptides to form functional protein molecules has been observed without exception. This is true even of polypeptides containing large numbers of cysteine residues, which must fold to form a precise pattern of disulfide bonds, frequently regarded as a daunting task [36]. Because large amounts (>100 milligrams) of high purity polypeptide are obtained in a single step from chemical ligation of unprotected peptide segments [37], it is possible to use several one milligram samples to undertake a preliminary examination of a range of typical folding conditions in a matter of only a few hours. Optimal conditions can then be used for folding of the bulk of the synthetic material. An example of the folding of a small, cysteine-rich chemically synthesized protein using this approach is shown in Figure 4.

It may be that the invariably efficient folding of chemically synthesized polypeptides is simply a matter of starting with abundant amounts of material of high chemical purity. Alternatively, the experimentally observed facile folding of proteins prepared by total chemical synthesis may be due to the fact that the preparation of chemically synthesized polypeptides does not proceed via an intermediate 'inclusion body' (i.e. aggregated) state.

Analytical characterization

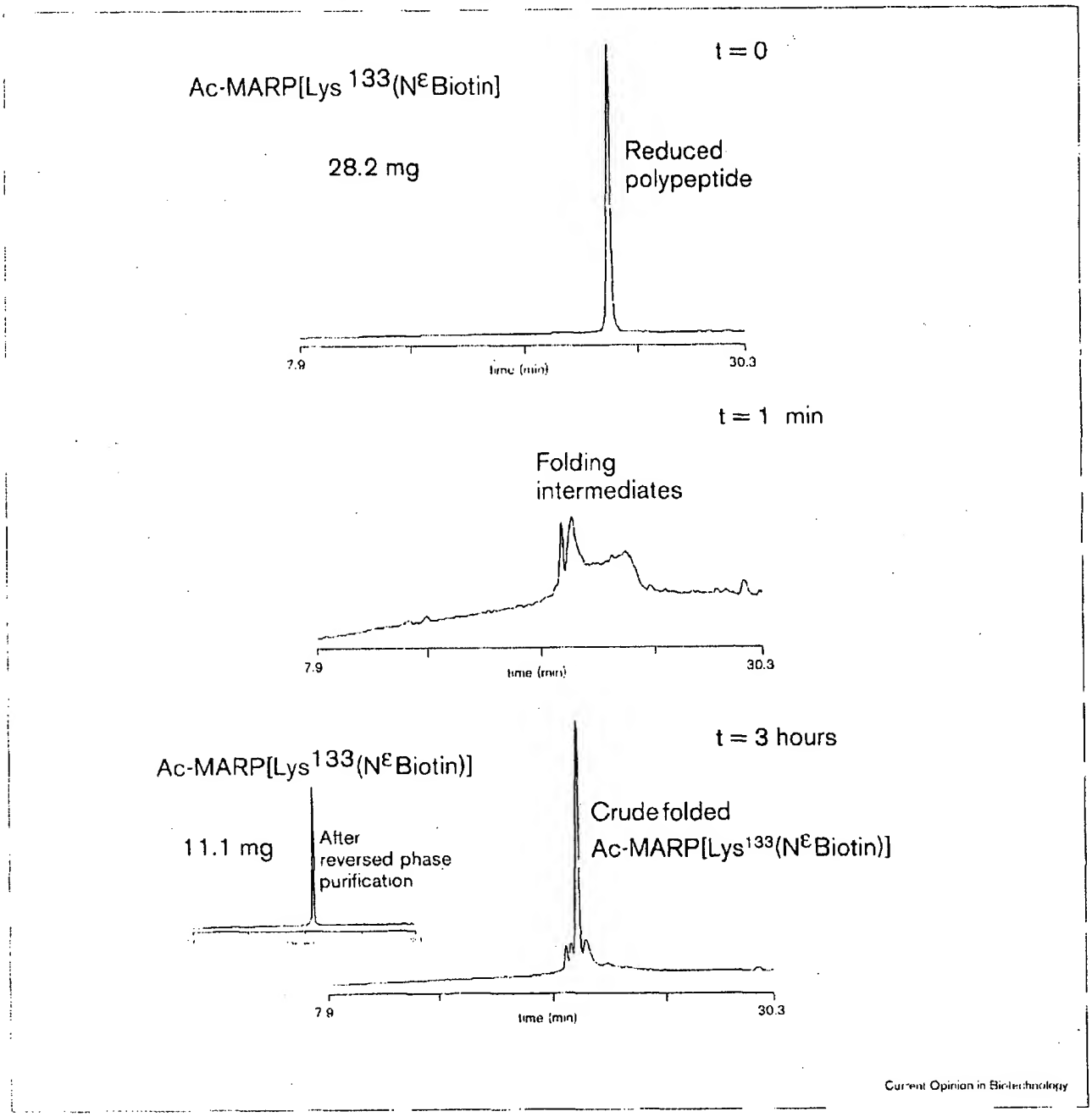
Synthetic protein chemists of the late 20th century have the advantage of powerful analytical techniques for determining the chemical structure of the protein molecules that they construct.

The covalent structure of synthetic protein molecules can be rapidly and accurately determined by electrospray ionization mass spectrometry (ESMS) [38]. Unprotected polypeptides, regardless of size, are efficiently analyzed in this way. In just a few minutes, the mass of a polypeptide can be measured with an accuracy of better than 1 part in 10,000. For the approximately fifty-residue unprotected peptide segments used in protein synthesis by chemical ligation, this means that the mass of each segment can be determined to better than ± 0.5 Da. Such accurate mass measurement, together with the knowledge of the amino acid sequence and the synthetic procedures used, is powerful confirmation of the correct chemical structure of the peptide segments used as building blocks in protein synthesis by chemical ligation. ESMS also yields useful data concerning purity of polypeptides [39]. This ESMS data complements and extends that obtained by more traditional analytical techniques, such as reverse phase HPLC [40]. Ligated polypeptide products are also characterized by ESMS with the same precision as the peptide segments [28•,29••,30,37]. Furthermore, for synthetic functional domains of typical size, the mass accuracy is sufficient to determine that the correct number of disulfide bonds has formed in the folding. For example, the folding of the 67 residue ligated synthetic polypeptide (measured mass 7791.6 Da) to form the chemokine SDF-1α (measured mass 7787.2 Da) resulted in the loss of 4.4 Da, precisely corresponding to formation of two disulfide bonds [31•]. In another example of the use of mass spectrometry to evaluate correct protein folding, ESMS was used to characterize the insertion of metals into a zinc-finger protein domain (Figure 5).

The three-dimensional structure of the folded synthetic protein can also be rapidly evaluated using modern multidimensional NMR techniques. The standard research scale of chemical protein synthesis gives multiple tens-of-milligrams of high purity synthetic protein directly from sequence data. Such amounts are ideal for NMR studies. Simple multi-dimensional 'fingerprint' measurements will quickly tell whether a synthetic protein has a unique folded form [41], or consists of a mixture of forms [42]. An example of such a measurement on the protein AOP-RANTES, prepared by total chemical synthesis, is shown in Figure 6. More complete NMR measurements on a synthetic protein can be used to experimentally determine the complete three dimensional structure of novel molecules [43].

X-ray crystallography can also be used to determine the three-dimensional structures of novel proteins prepared by

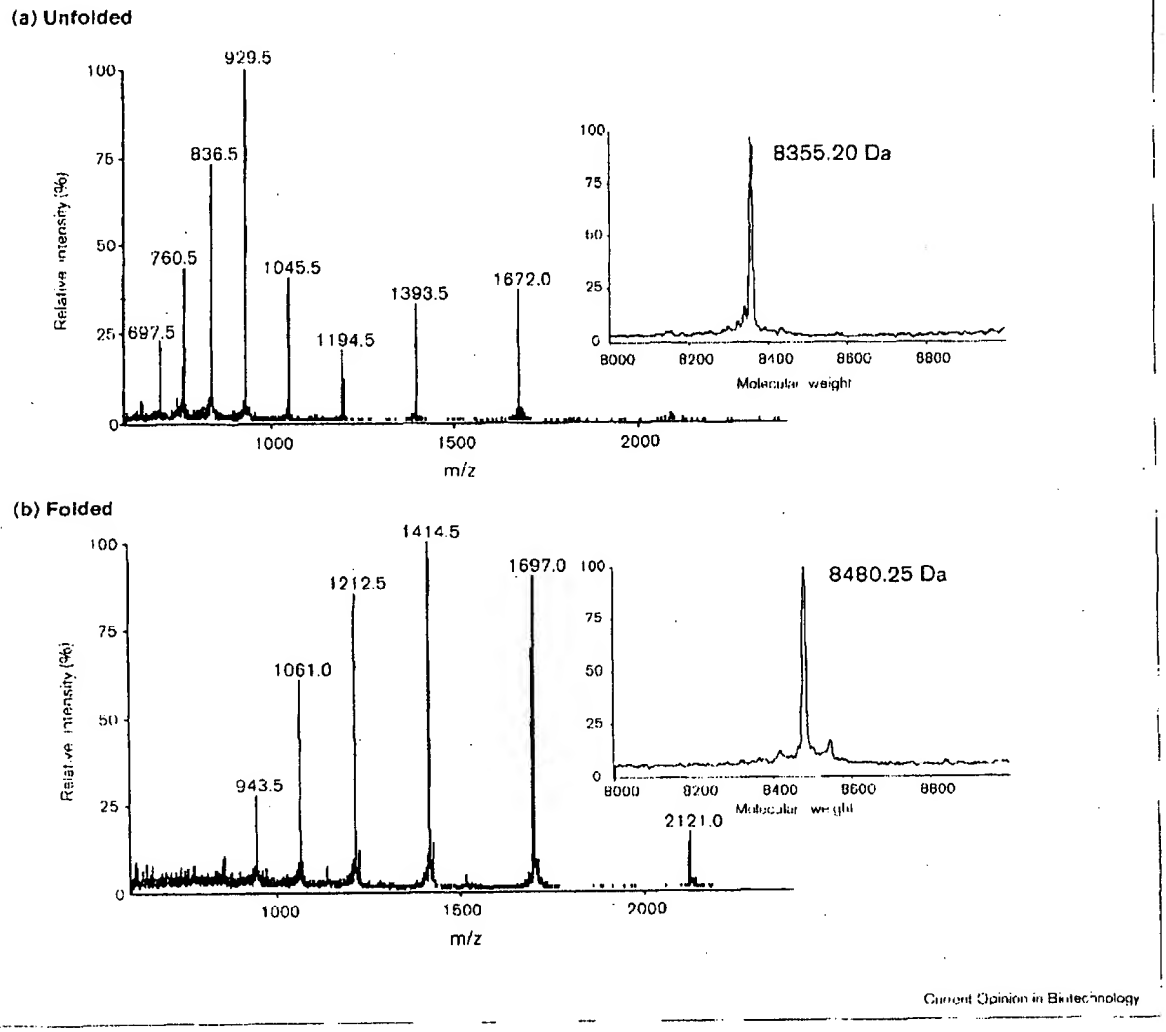
Figure 4



Efficient folding of a chemically synthesized protein. The cysteine-rich carboxy-terminal domain of the Agouti-related protein, termed 'minimized Agouti-related protein' (MARP), has 46 amino acids and includes ten cysteine residues. Folding of such small cysteine-rich polypeptides is regarded as difficult [36]. As can be seen here, the synthetic polypeptide containing a site-specific biotin label, folds efficiently to give a good yield of high purity labeled protein (at lower left; 'Ac-MARP[Lys^{133}(N$^{\varepsilon}$Biotin)'). Multidimensional NMR measurements show that this synthetic protein has a single defined three-dimensional structure, and bioassays show that the protein retains all the activities of the Agouti-related protein (DA Thompson, personal communication).

total chemical synthesis. A number of proteins prepared by total chemical synthesis based on genome sequence data have been crystallized and shown to diffract X-rays to high resolution [10,11,25••,44]. The crystal structure of a chemically synthesized protein is shown in Figure 7. The incorporation of [4Se]Met into a protein by chemical synthesis, together with use of a synchrotron radiation source for the diffraction studies, provides phasing information that can be used to solve a structure without making multiple isomorphous heavy atom crystalline forms. Initiatives

Figure 5



Insertion of metals into the chemically synthesized zinc finger protein DSX, the DNA-binding domain of the sex-determining gene product in *Drosophila*, observed by electrospray mass spectrometry. **(a)** Unfolded DSX synthetic 72 amino acid polypeptide: observed molecular mass 8355.2 ± 0.25 Da (calculated 8353.8 Da. average isotope composition). **(b)** Folded DSX $(Zn^{2+})_2$ protein: observed molecular mass 8480.3 ± 0.4 Da. The mass difference between the unfolded synthetic 72 residue polypeptide chain and the folded, metal-containing protein was 125.1 ± 0.7 Da in good agreement with four ionized cysteine side chains binding two Zn ions (i.e. 2 × 65.3 Da − 4 × 1 Da = 126.6 Da). Note the change in charge state distribution characteristic of proteins ionized from the native, folded state.

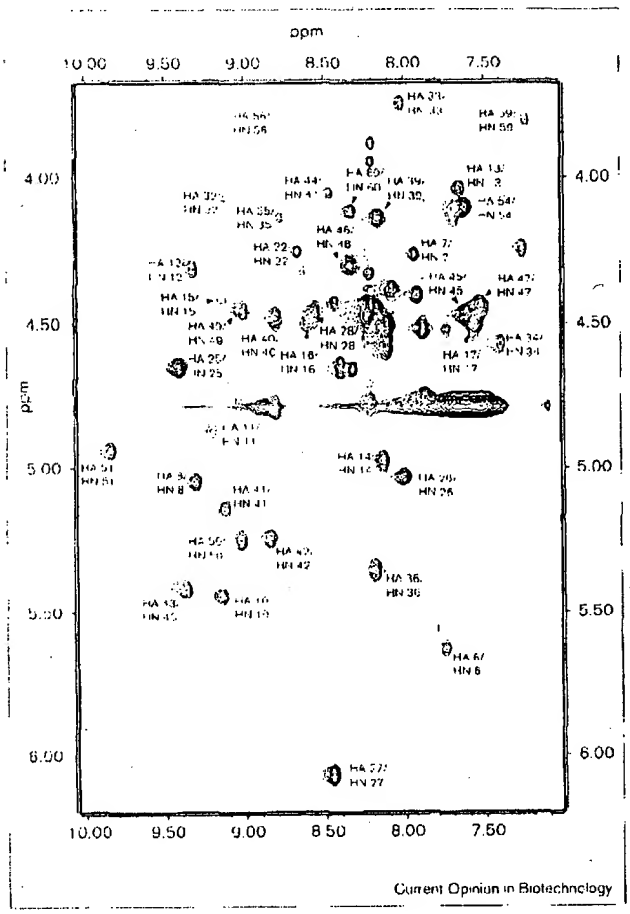
are under way to use X-ray crystallography and chemical protein synthesis to obtain experimental structures for entire families of molecules.

Functional characterization

Proteins prepared by total chemical synthesis are routinely assayed for biological and biochemical activity. After folding, purified synthetic proteins have been shown to have activities indistinguishable from reference materials, typically proteins purified from recDNA-expression systems [31•]. More and more frequently, however, the 'cloned' protein is not available in a timely fashion and the chemically synthesized protein becomes the reference material for that field of research [45,46].

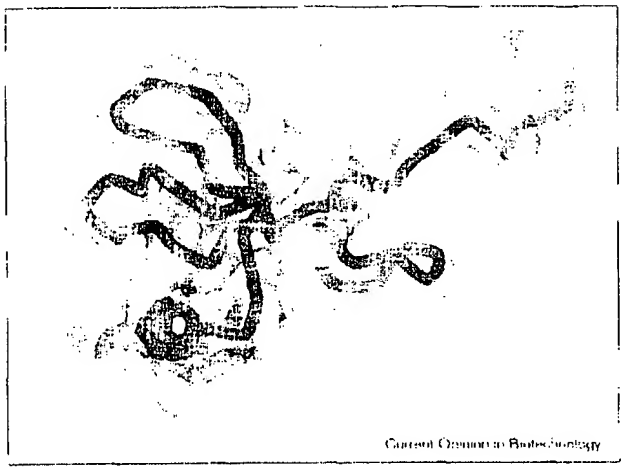
A good example of this is the chemokine family of proteins. The chemokines (the name is a contraction of '*chemo*tractant cyto*kines*') are a family of small (~70 amino acid) proteins that have a conserved pattern of cysteine residues (usually four). They are involved in all aspects of human inflammatory and immune responses and have other activities, such as hematopoietic effects on bone marrow cells. Recently, chemokine receptors have been shown to be the essential co-receptors for HIV, and the chemokines

Figure 6



^{1}H NMR measurements on the synthetic protein AOP-RANTES. Data shown is a 2D TOCSY measurement performed at 500 MHz. The data defines a single defined 3D structure, consistent with the canonical chemokine fold. Full assignment and interpretation of a complete set on NMR measurements is under way (PN Barlow, H McSparron, DA Thompson and J Wilken, personal communication).

Figure 7



Three-dimensional structure of the chemokine [N33A]SDF-1α determined by X-ray diffraction studies performed on crystalline chemically synthesized protein [44]. The polypeptide backbone is shown, together with the space filling van der Waals contact surface of the protein molecule.

themselves can block the entry of HIV-1 into human cells. For these reasons, chemokines are of great interest to the biomedical research community [47]. Furthermore, because of their small size and distinctive pattern of cysteine residues, chemokines have been identified in increasing numbers in the expressed sequence tag nucleic acid sequence databases. More than 40 human chemokine proteins have now been identified, together with a growing number of viral chemokines active against human receptors.

Chemical protein synthesis has been used to elucidate much of the original knowledge of the biology of the chemokines [17]. A case in point is the chemokine SDF-1α, the natural ligand for the CXCR4 receptor. CXCR4, formerly known as Lester or Fusin, is the coreceptor for HIV-1 entry into human T cells. SDF-1α was identified as the natural ligand for CXCR4 using chemokine prepared by total chemical synthesis [17,45,46]; most follow up studies on SDF-1α have also been undertaken using material prepared by chemical protein synthesis [48]. When material from recDNA-based expression in *E. coli* finally became available, it was shown to have activity indistinguishable from the reference synthetic SDF-1α prepared by native chemical ligation (Figure 8) [31•].

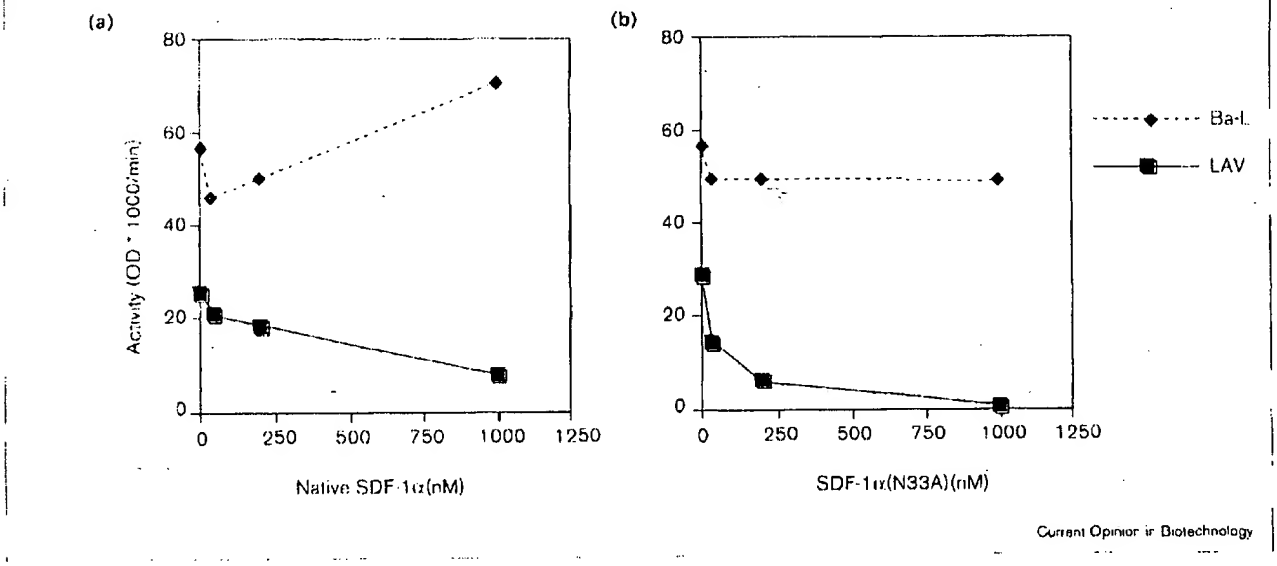
Multidomain proteins

Based on abundant experimental evidence, there can be no doubt that synthetic polypeptide chains corresponding to individual protein domains fold efficiently and correctly. These domains are the basic building blocks of the protein world. The demonstration that a wide range of fully functional domains can be efficiently formed from chemically synthesized polypeptide chains constitutes strong evidence that the world of protein function is generally accessible to the synthetic chemistry approach. Admittedly, thus far only a limited number of multidomain proteins have been prepared by total chemical synthesis (see Tables 1 and 2). Nonetheless, we can be absolutely confident of general access to large multidomain proteins because, if necessary, the individual domains can first be folded, then stitched together as folded protein molecules using established chemical techniques, such as oxime-forming ligation, first pioneered with synthetic manipulations of recombinant proteins [49].

Applications of chemical protein synthesis

Total synthesis provides for the straightforward application of the tools of chemistry to the world of proteins. This has a number of immediate and useful consequences. Molecular constructs that are highly desirable but that previously have been virtually impossible to make, even using elaborate novel recDNA-based *in vitro* expression

Figure 8



Biological assay data for [N33A]SDF-1α versus recHumanSDF1-α. Inhibition of CXCR4-mediated HIV-1-induced cell fusion by increasing concentrations of the chemokine SDF-1α. (a) Recombinant human protein. (b) Chemically synthesized [N33A]SDF-1α. Two different HIV envelopes were used: Ba-L uses the chemokine receptor CCR5 as an entry co-factor; LAV uses CXCR4. The natural ligand for CXCR4 is the chemokine SDF-1α, thus increasing concentrations of SDF1-α inhibit LAV-mediated fusion, but have no effect on Ba-L mediated fusion. (M Siani, E Berger, personal communication).

technologies [50], become accessible in a direct and practical fashion by chemical protein synthesis.

Non-coded amino acids

Non coded amino acids can be readily incorporated into proteins by total chemical synthesis. An early example of the power of this approach was the [$Aba^{67,95,167,195}$]HIV-1 protease molecule prepared by total chemical synthesis [9] and used for the original X-ray crystallography of this important target for drug design [10,11] (see above). The four cysteine residues in the enzyme molecule were replaced by L-α-NH_2-*n*-butyric acid (Aba) residues, effectively replacing the side chain -SH moiety of each cysteine residue with the isosteric -CH_3 group. By removing unpaired cysteine sulfhydryl groups, this substitution greatly improved the handling properties of the protein without affecting its structure and enzymatic properties [9–11]. Many subsequent examples of the incorporation of non-coded amino acids by total chemical protein synthesis have been reported, some involving more elaborate amino acid replacements, such as the incorporation of a geometrically constrained β-turn dipeptide mimic into a functional protein [51]. Similarly, constrained helical peptidomimetics have been reported [52] and could also be used as building blocks for the chemical construction of proteins.

Protein structure-activity relationships

Chemical protein synthesis has been used to carry out intimate studies elucidating the structural origins of protein function. For example, Love *et al.* [53•], prepared a novel form of the protein ubiquitin, in which two amino acids in the hydrophobic core were replaced by non-coded residues: Val26→Aba; and Ile30→Nva (L-*nor*Valine) (Aba being straight chain aliphatic [-CH_2CH_3] and Nva being straight chain aliphatic [-$CH_2CH_2CH_3$]). In each case, these changes amount to the deletion of a single methyl group from the native sidechain. The synthetic protein variant was crystallized and X-ray diffraction showed an overall contraction of the molecule, and the introduction of a small internal cavity, which decreased the stability of the protein. The modified protein showed unaltered biological activities [53•].

Protein backbone engineering

The systematic modification of the chemical nature of the polypeptide backbone structure, and the correlation of such changes with effects on protein function, is also made uniquely possible by chemical protein synthesis. Baca and Kent [54] used deletion of H-bond donor moieties (by conversion of -CONH- to -COS-) in HIV-1 protease prepared by chemical ligation to show that protein 'flap' backbone-to-substrate hydrogen bonds, mediated by a tetrahedrally coordinated water molecule, were essential to full catalytic activity of this enzyme. More recently, Baca and co-workers [25••] extended and refined this work by using convergent chemical ligation to construct a 22 kDa covalent tethered dimer form of the HIV-1 protease [23] in which the backbone H-bonding element of only a single

flap was deleted. This backbone-engineered HIV-1 protease retained full intrinsic catalytic activity [25••]. Such studies suggested that the HIV-1 protease enzyme molecule, a homodimer of identical 99 residue peptide chains constituting a single active site, uses only a single flap in catalysis, in a manner analogous to the single polypeptide chain, two domain cell-encoded aspartyl protease enzymes [25••]. This finding has significant implications for the correct understanding of the mechanism of the HIV-1 protease, and for the design of improved protease inhibitor AIDS therapeutics.

Intermolecular hydrogen bonding

Backbone to-backbone amide hydrogen bonds are frequently observed in protein complexes. The contribution of intermolecular backbone hydrogen bonding to protein–protein interactions was experimentally quantified by specific deletion of a single H-bonding element (i.e. conversion of a backbone -CONH- to -COO-) in synthetic turkey ovomucoid third domain, a small serine protease inhibitor protein [55•]. Evaluation of the binding of this backbone ester-containing protein inhibitor molecule to a panel of serine proteinases showed that this hydrogen bond contributed 1.5 ± 0.3 kcal/mole to the formation of the enzyme–ligand complex. Such studies are made uniquely possible by chemical protein synthesis.

Mirror image D-proteins

Total chemical synthesis can be used to prepare polypeptide chains made up entirely of D-amino acids, rather than the L-amino acid polypeptide chains found in nature as the products of ribosomal synthesis. Such D-amino acid polypeptides fold with high efficiency to form proteins of defined three-dimensional structure; the resulting protein molecules are the mirror image of the native L-proteins, and display reciprocal chiral properties (i.e. circular dichroism spectra; mirror image folds by NMR and by X-ray crystallography; mirror image substrate chiral specificities in enzymes). An early example of the total synthesis of mirror image proteins was the D- and L-forms of the trypsin inhibitor microprotein EETI II [56]. The protein enantiomers D- and L-rubredoxin prepared by stepwise SPPS [57], and D- and L-HIV-1 protease prepared by chemical ligation [21], were both studied by X-ray crystallography. The three-dimensional structure of ligated D-HIV-1 protease and the mechanistic significance of this work and other chemical variants of this important enzyme are discussed by Miller *et al.* [25••]. More recently, Schumacher *et al.* [58] used mirror image SH3 domains, prepared by chemical synthesis, to identify novel ligands in phage peptide libraries by 'mirror-image phage display'. There is also a recent preliminary report of crystallography on the mirror image forms of the sweet protein monellin [59]. Mirror image proteins can be used to obtain crystal structures of greater accuracy [57], and to discover novel lead compounds from chiral libraries [58]. Although mirror image proteins could be expected to have improved properties *in vivo* (e.g. resistance to proteases, little or no immunogenicity), only D-proteins that work on achiral ligands (e.g. superoxide dismutase) will retain their biological activity.

Glycoproteins

The chemoselective reaction principle, in which unprotected macromolecular building blocks are set up to react with one another to yield defined ligation products of unique structure, also can be applied to the total synthesis of glycoproteins. Zhao *et al.* [60] linked complex carbohydrates to synthetic peptides by reaction of the aldehyde or ketone moiety at the reducing end of the carbohydrate with an NH_2OCH_2- moiety synthetically introduced on the peptide, to give the corresponding homogeneous oxime-linked glycopeptide of defined molecular composition. Bertozzi and co-workers [61•,62•] used chemoselective ligation within the carbohydrate structure to elaborate synthetic glycopeptides, and have applied this approach to elegant glycoprotein synthesis. Chemoselective synthesis of glycoproteins of defined homogeneous molecular structure would have profound biotechnology applications in the field of human biopharmaceuticals. The current state of chemical approaches to glycobiology has been reviewed [63].

Circular proteins

Topologically circular (i.e. cyclic) polypeptides that fold to form proteins are a novel class of molecules uniquely accessible by chemistry. Muir and co-workers [64••] reported the synthesis of a topologically circular polypeptide chain corresponding to the WW domain of the intracellular signaling protein Yes kinase-associated protein (YAP). The 46 residue polypeptide chain was converted to cyclic form (i.e. a topologically circular polypeptide, with no amino-terminus, and no carboxy-terminus) using the native chemical ligation method [27] as applied to cyclic peptides [65,66]. Interestingly, the circular polypeptide folded efficiently to form a protein molecule with activity and three-dimensional structure indistinguishable from the native (linear polypeptide) protein. Circular proteins could have great potential application in biotechnology: they would be expected to display increased thermal stability and to resist the action of exoproteases. Fundamental insights into protein folding will be obtained by the study of these protein molecules formed from polypeptides with circular topology.

De novo design

Chemical synthesis can be used to construct artificial proteins based on *de novo* molecular design. The principles underlying this chemoselective reaction approach were elegantly demonstrated by Rose in 1994 [67]. More recently, a number of examples have been reported of the total synthesis of TASP (Template Assisted Synthetic Protein) molecules, using the modular chemical ligation approach first applied to TASPs by Dawson [68] and Rose [67]. These include the construction of a 14.3 kDa heme-containing TASP [69•]. Practical considerations governing TASP design and construction by chemical ligation have also been investigated [70].

DeGrado [71] recently summarized the current state of the art in the *de novo* design of protein molecules. Significant progress has been made on several fronts, including Regan and co-workers' [72] protein structural redesign (systematic changes of amino acid sequence causing a protein to form a different predetermined fold), and the use of design principles to equip proteins with new functions [73]. True *de novo* design of novel polypeptide sequences encoding a predetermined protein fold has also advanced considerably, and Dahiyat and Mayo [74•] have tested the success of their novel predictive algorithms by chemical synthesis and structural characterization of the designed molecules. Chemical protein synthesis can play a pivotal role in the practical evaluation of protein design because it provides a direct route to the protein molecule without the intervention of the extraneous and potentially complicating factors found in living systems [33].

Conclusions and future perspectives

The literature in the past year or so is characterized by a growing number of examples of the study of the properties of a variety of small proteins produced by total chemical synthesis [75–80]. This is evidence that the field of chemical protein synthesis is moving from the pioneering 'feasibility' stage to the use of chemically synthesized proteins for serious study of protein function. Total synthesis using the chemical ligation of unprotected peptide segments in aqueous solution has already provided robust and practical synthetic access to proteins in the 7,000–22,000 Da size range. Improvements (see below) promise to dramatically expand the type and size of molecules which can be studied by the application of chemistry to the world of proteins.

Novel methodologies

Solid phase chemical protein synthesis

Canne *et al.* [81] have provided a preliminary report that describes the extension of the chemical ligation principle to polymer-supported synthesis. 'Solid phase chemical ligation', the chemical ligation of unprotected peptides on a polymer support, takes advantage of rapid purification by filtration of intermediate polymer-bound ligation products and, thus, makes feasible the sequential ligation of up to eight peptide segments. The principles of the approach are shown in Figure 9. Completely unprotected peptide segments were condensed in aqueous solution on a water-compatible solid phase. The method was illustrated with the synthesis of the 115 amino acid polypeptide of macrophage migration inhibitory factor by sequential ligation of four unprotected peptide segments [81]. Even at this early stage of its development, solid phase chemical ligation promises to provide facile and robust synthetic access to polypeptides in the 120–200 amino acid size range. This corresponds to the size range of protein 'domains', the autonomous units of folding and function that are the building blocks of the protein world [82].

Expressed protein ligation

Muir and co-workers' [83••,84] have reported a recDNA-based technology that, together with the native chemical ligation method in its original form [27], promises to fully integrate the worlds of recDNA-based protein expression and chemical peptide synthesis. It has been possible since the first publication of the native chemical ligation method [27] to covalently join synthetic peptides to proteins expressed with an amino-terminal cysteine residue. The newer 'expressed protein ligation' method accomplishes the inverse, allowing the thioester-mediated native chemical ligation of expressed polypeptides to synthetic peptides containing an amino-terminal cysteine residue [83••,84]. It takes advantage of an ingenious defective intein expression vector that generates a transient thioester in a recombinantly expressed protein. In this way, synthetic peptides containing amino-terminal cysteine residues can be ligated with reasonable efficiency to the carboxy-terminal of proteins produced by standard recDNA-expression in microorganisms. Mechanistic studies indicate that expressed protein ligation proceeds according to the thioester-mediated native chemical ligation mechanism [83••]. Together with the Verdine laboratory's application of the original native chemical ligation method [85], expressed protein ligation makes possible the introduction of precise chemical modifications to a range of cloned/expressed proteins.

Novel applications

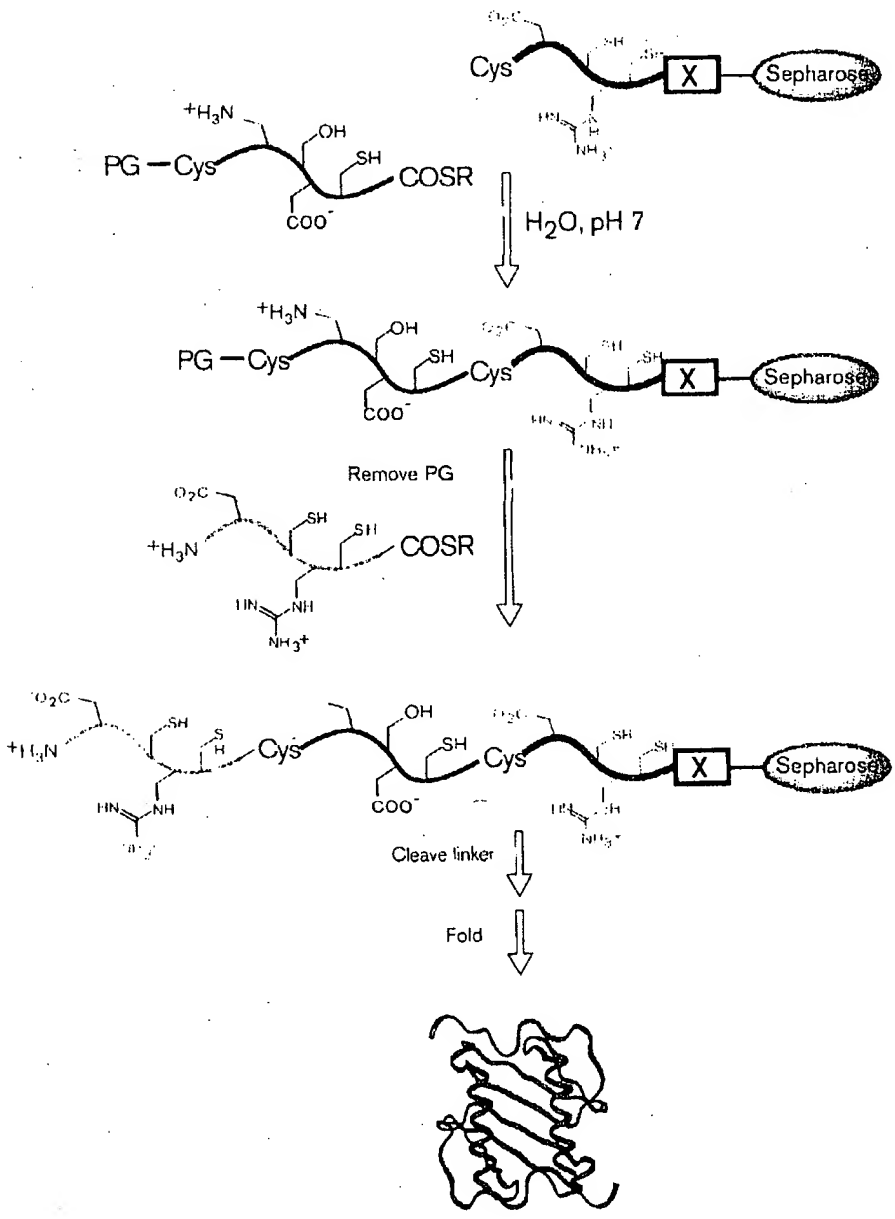
Protein signature analysis

The principles of combinatorial chemistry have also been applied to protein molecules [86]. The protein signature analysis method involves the total synthesis of self-encoded libraries of protein analogs [87•] followed by a selection step, such as affinity chromatography, to separate functional analogues from the pool of non-functional molecules. The pools are decoded in one step using a chemical self-encoding scheme applied to proteins [86,87•]. In this way, a single experiment generates patterns of information that describe the structural origins of function in the protein molecule. As chemical access to the world of proteins grows more robust, it can be expected that there will be further examples of the application of the principles of combinatorial chemistry to the world of proteins.

Precision labeling

In practical terms, one of the most useful future applications of total chemical synthesis will be the site-specific labeling of proteins with reporter entities for physical measurements and for precision assays. Chemical protein synthesis allows proteins to be precisely labeled at predetermined sites with complete versatility. Using chemistry, it is as straightforward to introduce any of a variety of labels at one or more specific sites in a synthetic protein as it is to introduce a standard amino acid. Examples include the site-specific incorporation of biotin (Figure 4), and the synthesis of a variety of fluorescently-labeled proteins, each consisting of a single molecular species. Such precision-labeled proteins, not readily obtainable using

Figure 9



Principles of solid phase chemical ligation. The carboxy-terminal unprotected peptide segment is attached by a cleavable linker moiety to a water-compatible ploymer support. Sidechain-unprotected peptide-thioester segments are sequentially reacted by native chemical ligation, to yield the full length polymer bound polypeptide. The polypeptide product is released from the support, purified and folded to give the target protein.

recDNA techniques, are invaluable reagents for a variety of biological studies and for assay development.

Protein families

Another useful application of chemical protein synthesis is the rapid preparation of families of proteins directly from sequence data. Chemical incorporation of NMR probe nuclei or modified amino acids with suitable X-ray scattering properties, such as [4Se]Met, enables the rapid elucidation of protein structures.

Membrane proteins

Finally, the application of chemical protein synthesis to provide direct access to integral membrane proteins is an

intriguing possibility. Membrane proteins constitute 20–30% of known proteins across a variety of genomes, based on analysis of open reading frames [88]. They function, among other things, as ion channels, transporter systems, and receptor molecules. Very little is known about the structural basis of function for this important class of proteins because, to date, integral membrane proteins have resisted attempts by recDNA-based expression to produce the large amounts of high purity protein required for high resolution structure determination. It may be that total protein synthesis by chemical ligation will provide alternative strategies for the production of large amounts (tens-to-hundreds of milligrams) of the integral membrane proteins. This would have a profound effect on biological research.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Frisch C, Schreiber G, Johnson CM, Fersht AR: **Thermodynamics of the interaction of barnase and barstar: changes in free energy versus changes in enthalpy on mutation.** *J Mol Biol* 1997, **267**:696-706.

2. Winter G, Fersht AR, Wilkinson AJ, Zoller M, Smith M: **Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding.** *Nature* 1982, **299**:756-758.

3. Fisher E: **Untersuchungen uber aminosauren, polypeptide, und proteine.** *Ber Chem Ges* 1906, **39**:530-610.

4. Seiber P, Kamber B, Hartmann A, Johl A, Riniker B, Rittel W: **Total synthese von human insulin unter gezielter bildung der disulfidbindungen.** *Helv Chem Acta* 1974, **57**:2617-2621.

5. Sakakibara S: **Synthesis of large peptides in solution.** *Biopolymers (Pept Sci)* 1995, **37**:17-28.

6. Inui T, Bodi J, Kubo S, Hishio H, Kimura T, Kojima S, Maruta H, Muramatsu T, Sakakibara S: **Solution synthesis of human midkine, a novel heparin-binding neurotrophic factor consisting of 121 amino acid residues with five disulphide bonds.** *J Pept Sci* 1996, 2:28-39.

7. Kojima S, Inui T, Muramatsu H, Suzuki Y, Kadomatsu K, Yoshizawa M, Hirose S, Kimura T, Sakakibara S, Muramatsu T: **Dimerization of midkine by tissue transglutaminase and its functional implication.** *J Biol Chem* 1997, **272**:9410-9416.

8. Merrifield RB: **Solid phase peptide synthesis.** *Science* 1986, 232:341-347.

9. Schneider J, Kent SBH: **Enzymatic activity of a synthetic 99 residue protein corresponding to the putative HIV-1 protease.** *Cell* 1988, 54:363-368.

10. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J, Kent SBH: **Crystal structure of synthetic HIV-1 protease: conserved fold in retroviral proteases.** *Science* 1989, **245**:616-621.

11. Miller M, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Schneider J, Kent SBH, Wlodawer A: **Structure of a complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution.** *Science* 1989, **246**:1149-1152.

12. Miller M, Geller M, Gribskov M, Kent SBH: **Analysis of the structure of chemically synthesized HIV-1 protease complexed with a hexapeptide inhibitor. Part I: crystallographic refinement of 2.0 Angstrom data.** *Proteins - Struct Funct Genet* 1997, **27**:184-194.

13. Palella FJ Jr, Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA, Aschman DJ, Holmberg SD: **Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection.** *N Engl J Med* 1998, **338**:853-860.

14. Jackson DY, Burnier J, Quan C, Stanley M, Tom J, Wells JA: **A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues.** *Science* 1994, **266**:243-247.

15. Witte K, Sears P, Martin R, Wong C-H: **Enzymatic glycoprotein synthesis: preparation of ribonuclease glycoforms via enzymatic glycopeptide condensation and glycosylation.** *J Am Chem Soc* 1997, **119**:2114-2118.
• Methods are described for the chemoenzymatic synthesis of defined carbohydrate structures, both natural and unnatural. The potential of this approach for the total synthesis of glycoproteins is illustrated by the enzymatic re-ligation of peptide fragments to give a series of RNase B glycoforms.

16. Tatusov RL, Koonin EV, Lipman DJ: **A genomic perspective on protein families.** In *Genome Issue: Building Gene Families. Science* 1997, **278**:631-637.

17. Clark-Lewis I, Vo L, Owen P, Anderson J: **Chemical synthesis, purification, and folding of C-X-C and C-C chemokines.** *Methods Enzymol* 1997, **287**:233-250.

18. Rajarathnam K, Kay CM, Clark-Lewis I, Sykes BD: **Characterization of quaternary structure of interleukin-8 and functional implications.** *Methods Enzymol* 1997, **287**:89-105.
•• This article is a definitive summary of the imaginative application of total chemical protein synthesis to elucidate the molecular basis of function in the prototypical chemokine interleukin-8. Of particular note is the use of chemical backbone modification to define the full biological activity of the monomeric form of the chemokine.

19. Rajarathnam K, Sykes BD, Kay CM, Dewald B, Geiser T, Baggiolini M, Clark-Lewis I: **Neutrophil activation by monomeric interleukin 8.** *Science* 1994, **264**:90-92.

20. Schnölzer M, Kent SBH: **Constructing proteins by dovetailing unprotected synthetic peptides: backbone engineered HIV protease.** *Science* 1992, **256**:221-225.

21. deLisle Milton R, Milton S, Schnolzer M, Kent SBH: **Synthesis of proteins by chemical ligation of unprotected peptide segments: mirror-image enzyme molecules, D- & L-HIV protease analogues.** In *Techniques in Protein Chemistry IV*. Edited by Angeletti R. New York: Academic Press; 1993:257-267.

22. Schnölzer M, Alewood P, Alewood D, Kent SBH: **"In situ" neutralization protocols in Boc-chemistry solid phase peptide synthesis: rapid, high yield assembly of difficult sequences.** *Int J Pept Protein Res* 1992, **40**:180-193.

23. Baca M, Muir TW, Schnölzer M, Kent SBH: **Chemical ligation of cysteine-containing peptides: synthesis of a 22kDa tethered dimer of HIV-1 protease.** *J Am Chem Soc* 1995, **117**:1881-1887.

24. Williams M, Muir TW, Ginsberg M, Kent SBH: **Total chemical synthesis of a folded β-sandwich protein domain: a fibronectin type 3 module.** *J Am Chem Soc* 1994, **116**:10797-10798.

25. Miller M, Baca M, Rao JKM, Kent SBH: **Probing the structural basis of the catalytic activity of HIV-1 protease through total chemical protein synthesis.** *J Mol Struct* 1998, **423**:137-152.
•• This article summarizes a series of investigations of the HIV-1 protease by total chemical synthesis. It includes the first report of the high resolution crystal structure of a mirror-image enzyme molecule, and the total synthesis of a backbone-engineered fully active covalent tethered-dimer form of the HIV-1 protease. Both these unusual molecular forms display features incompatible with accepted molecular mechanisms for catalysis and inhibition of this enzyme.

26. Canne LE, Ferré-D'Amare AR, Burley SK, Kent SBH: **Total chemical synthesis of a unique transcription factor-related protein: cMyc-Max.** *J Am Chem Soc* 1995, **117**:2998-3007.

27. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH: **Synthesis of proteins by native chemical ligation.** *Science* 1994, **266**:776-779.

28. Hackeng TM, Mounier CM, Bon C, Dawson PE, Griffin JH, Kent SBH: **Total chemical synthesis of enzymatically active human type II secretory phospholipase A2.** *Proc Natl Acad Sci USA* 1997, **94**:7845-7850.
• Describes the total chemical synthesis of the 124 residue polypeptide with 14 cysteine residues. The molecule folded correctly to give a protein containing seven disulfides and displaying full enzymatic activity.

29. Dawson PE, Churchill M, Ghadiri MR, Kent SBH: **Modulation of reactivity in native chemical ligation through the use of thiol additives.** *J Am Chem Soc* 1997, **119**:4325-4329.
•• Definitive elucidation of the mechanism of native chemical ligation. The mechanistic origins of the exquisite regioselectivity of the chemoselective ligation at Cys residues are explained. Total syntheses of analogues of the enzyme 'barnase' (110 residues) are used as model studies.

30. Muir TW, Dawson PE, Kent SBH: **Protein synthesis by chemical ligation of unprotected peptide segments in aqueous solution.** In

Methods in Enzymology: Solid Phase Peptide Synthesis, vol 289. Edited by Fields GB. New York: Academic Press; 1997:266-298.

31. • Ueda H, Siani MA, Gong W, Thompson DA, Brown GG, Wang JM: **Chemically synthesized SDF-1 analogue, N33A, is a potent chemotactic agent for CXCR4/Fusin/LESTR-expressing human leukocytes.** *J Biol Chem* 1997, **272**:24966-24960.

A meticulous comparison of the binding and chemotactic activities of SDF-1α preparations from recombinant expression in *E coli* and from total synthesis by native chemical ligation. Protein from the two sources, biosynthetic and chemical, had indistinguishable activities.

32. Anfinsen CB: **Principles that govern the folding of protein chains.** *Science* 1973, **181**:223-230.

33. Todd MJ, Viitanen PV, Lorimer GH: **Dynamics of the chaperonin ATPase cycle: implications for facilitated protein folding.** *Science* 1994, **265**:659-66.

34. Mitraki A, King J: **Protein folding intermediates and inclusion body formation.** *Biotechnology* 1989, **7**:690-697.

35. Jaenicke R, Rudolph R: **Folding proteins.** In *Protein Structure: A Practical Approach*. Edited by Creighton TE. Oxford: IRL Press, Oxford University Press; 1989:191-223.

36. Fersht A: Cited in 'Fold me deadly'. *Nat Struct Biol* 1998, **5**:327-328.

37. Liu W, Qasim MA, Kent SBH: **Comparative total synthesis of turkey ovomucoid third domain by both stepwise solid phase peptide synthesis and native chemical ligation.** *J Am Chem Soc* 1996, **118**:8518-8523.

38. Chait BT, Kent SBH: **Weighing naked proteins: practical, high-accuracy mass measurement of peptides and proteins.** *Science* 1992, **257**:1885-1894.

39. Schnölzer M, Jones A, Alewood PF, Kent SBH: **Ionspray mass spectrometry in peptide synthesis: structural characterization of minor by-products in the synthesis of ACP(65-74).** *Anal Biochem* 1992, **204** 335-343

40. Fitzgerald MC, Kent SBH: **Total chemical synthesis of proteins.** In *Bioorganic Chemistry: Peptides and Proteins*. Edited by Hecht SM. New York: Oxford University Press; 1998:66-99.

41. Liu W, Starovasnik MA, Kent SBH: **Total chemical synthesis of bovine pancreatic trypsin inhibitor by native chemical ligation.** *FEBS Lett* 1998, **429**:31-35.

42. Muir TW, Williams M, Kent SBH: **Detection of synthetic protein isomers & conformers by electrospray mass spectrometry.** *Anal Biochem* 1995, **224**:100-109.

43. Crump MP, Gong JH, Loetscher P, Rajarathnam K, Amara A, Arenzana-Seisdedos F, Virelizier JL, Baggiolini M, Sykes BD, Clark-Lewis I: **Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 activation from binding and inhibition of HIV-1.** *EMBO J* 1997, **16**:6996-7007.

44. Dealwis C, Fernandez EJ, Thompson DA, Simon RJ, Siani MA, Lolis E: **Crystal structure of chemically synthesized [N33A]SDF-1α, a potent ligand for the HIV-1 'Fusin' co-receptor.** *Proc Natl Acad Sci USA* 1998, **95**:6941-6946.

45. Oberlin E, Amara A, Bachelerie F, Bessia C, Virelizier JL, Arenzana-Seisdedos F, Schwartz O, Heard JM, Clark-Lewis I, Legler DF *et al.*: **The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1.** *Nature* 1996, **382** 833-835.

46. Bleul CC, Farzan M, Choe H, Parolin C, Clark-Lewis I, Sodroski J, Springer TA: **The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry.** *Nature* 1996, **382**:829-833

47. Baggiolini M, Dewald B, Moser B: **Human chemokines: an update.** *Annu Rev Immunol* 1997, **15**:675-705

48. Campbell JJ, Hendrick J, Zlotnik A, Siani MA, Thompson DA, Butcher EC: **Chemokines and the arrest of lymphocytes rolling under flow conditions.** *Science* 1998, **279**:381-384.

49. Offord RE: **Going beyond the code.** *Protein Eng* 1991, **7**:709-710.

50. Hohsaka I, Ashizuka Y, Murakami H, Sisido M: **Incorporation of nonnatural amino acids into streptavidin through *in vitro* frame-shift suppression.** *J Am Chem Soc* 1996, **118**:9778-9779.

51. Baca M, Alewood PF, Kent SBH: **Structural engineering of HIV-1 protease with a β-turn mimic of fixed geometry.** *Protein Sci* 1993, **2**:1085-1091.

52. Phelan JC, Skelton NJ, Braisted A, McDowell RS: **A general method for constraining short peptides to an α-helical conformation.** *J Am Chem Soc* 1997, **119**:455-460.

53. • Love SG, Muir TW, Ramage R, Shaw KT, Alexeev D, Sawyer L, Kelly SM, Price NC, Arnold JE, Mee MP, Mayer RJ: **Synthetic, structural and biological studies of the ubiquitin system: synthesis and crystal structure of an analog containing unnatural amino acids.** *Biochem J* 1997, **323**:727-734.

Ingenious protein engineering, involving the deletion of single methyl groups from hydrophobic side chains in the core of the protein, possible only through total chemical protein synthesis.

54. Baca M, Kent SBH: **Catalytic contribution of flap-substrate hydrogen bonds in HIV-1 protease explored by chemical synthesis.** *Proc Natl Acad Sci USA* 1993, **90**:11638-11642.

55. • Lu W, Qasim MA, Laskowski M Jr, Kent SBH: **Probing intermolecular main chain H-bonding in serine proteinase–protein inhibitor complexes: chemical synthesis of backbone-engineered turkey ovomucoid third domain.** *Biochemistry* 1997, **36**:673-679.

Direct experimental assessment of the quantitative contribution of a backbone H-bond to protein–protein interaction; the answer is 1.5 ± 0.3 kcal/mole.

56. Nielsen KJ, Alewood D, Andrews J, Kent SBH, Craik DJ: **1H NMR determination of the three dimensional structures of mirror image forms of a Leu5 variant of the trypsin inhibitor Ecballium elaterium (EETI-II).** *Protein Sci* 1994, **3**:291-302.

57. Zawadzke LE, Berg JM: **The structure of a centrosymmetric protein crystal.** *Proteins* 1993, **16**:301-305.

58. Schumacher TNM, Mayr LM, Minor DL Jr, Milhollen MA, Burgess MW, Kim PS: **Identification of D-peptide ligands through mirror-image phage display.** *Science* 1996, **271**:1854-1857.

59. Hung L-W, Kohmura M, Ariyoshi Y, Kim S-H: **Crystallization and preliminary X-ray analysis of D-monellin.** *Acta Crystallogr Sect D Biol Crystallogr* 1997, **53**:327-328.

60. Zhao Y, Kent SBH, Chait BT: **Rapid, sensitive structure analysis of oligosaccharides.** *Proc Natl Acad Sci USA* 1997, **94**:1629-1633.

61. • Winans KA, King DS, Bertozzi CR: **Synthesis and structural characterization of an antibacterial glycoprotein.** CARB-086 in *Book of Abstracts. 214th ACS National Meeting*: 1997 September 7–11: Las Vegas, NV.

See annotation to [62•].

62. • Rodriguez EC, Winans KA, King DS, Bertozzi CR: **A strategy for the chemoselective synthesis of O-linked glycopeptides with native sugar-peptide linkages.** *J Am Chem Soc* 1997, **119**:9905-9906.

These reports from the Bertozzi laboratory at UC Berkeley highlight the application of chemoselective reaction to the elaboration of glycoforms, and illustrate the potential of this approach to the total synthesis of glycoproteins.

63. Yarema KJ, Bertozzi C: **Chemical approaches to glycobiology and emerging carbohydrate-based therapeutic agents.** *Curr Opin Chem Biol* 1998, **2**:49-61.

64. •• Camarero JA, Pavel J, Muir TW: **Chemical synthesis of a circular protein domain: evidence for folding-assisted cyclization.** *Angew Chem Int Ed* 1998, **37**:347-349.

Total chemical synthesis of a correctly folded, fully-functional protein that contains a topologically circular polypeptide chain. It is amazing that a peptide chain without amino- or carboxy-terminal residues (frequently invoked as nucleation sites for folding), and that has never been near the cell's folding machinery, folds quantitatively and correctly as demonstrated by two-dimensional NMR.

65. Zhang L, Tam JP: **Synthesis and application of unprotected cyclic peptides as building blocks for peptide dendrimers.** *J Am Chem Soc* 1997, **119**:2363-2370.

66. Camarero JA, Muir TW: **Chemoselective backbone cyclization of unprotected peptides.** *Chem Commun* 1997: 1369-1370.

67. Rose K: **Facile synthesis of homogeneous artificial proteins.** *J Am Chem Soc* 1994, **116**:30-33.

68. Dawson PE, Kent SBH: **Convenient total synthesis of a 4-Helix TASP molecule by chemoselective ligation.** *J Am Chem Soc* 1993, **115**:7263-7266.

69. • Rau HK, Haehnel W: **Design, synthesis, and properties of a novel Cytochrome b model.** *J Am Chem Soc* 1998, **120**:468-476.

The ultimate TASP synthesis, illustrating how this field has changed in the past four years: complex synthetic proteins of more than 14 kDa can be prepared in straightforward, unambiguous fashion by chemical ligation of unprotected peptide segments.

70. Wong AK, Jacobsen MP, Winzor DJ, Fairlie DP: **Template assembled synthetic proteins (TASPs): are template size, shape, and directionality important in formation of four-helix bundles?** *J Am Chem Soc* 1998, 120:3836-3841.

71. DeGrado WF: **Protein design enhanced: proteins from scratch.** *Science* 1997, 278:80-81.

72. Dalal S, Balasubramanian S, Regan L: **Protein alchemy: changing the beta sheet into alpha helix.** *Nat Struct Biol* 1997, 4:548-552.

73. Coldren CD, Hellinga HW, Caradonna JP: **The rational design and construction of a cuboidal iron-sulfur protein.** *Proc Nat Acad Sci USA* 1997, 94:6635-6640.

74 • Dahiyat BI, Mayo SL: ***De novo* protein design: fully automated sequence selection.** *Science* 1997, 278:82-87.
The objective selection of target sequences is an impressive step in the true *de novo* design of proteins. Also impressive is the and evaluation of novel proteins using chemical synthesis.

75. Klostermeier D, Bayer P, Kraft M, Frank RW, Roesch P: **Spectroscopic investigations of HIV-1 trans-activator and related peptides in aqueous solutions.** *Biophys Chem* 1997, 63:87-96.

76. Stenberg Y, Julenius K, Dahlqvist I, Drakenberg T, Stenflo J: **Calcium-binding properties of the third and fourth epidermal-growth-factor-like modules in vitamin-K-dependent protein.** *Eur J Biochem* 1997, 248:63-170.

77. De Rocquigny H, Petitjean P, Tanchou V, Decimo D, Drouot L, Delaunay T, Darlix J-L, Roques BP: **The zinc fingers of HIV nucleocapsid protein NCp7 direct interactions with the viral regulatory protein Vpr.** *J Biol Chem* 1997, 272:30753-30759.

78 Kuhlman B, Boice JA, Fairman R, Raleigh DP: **Structure and stability of the N-terminal domain of the ribosomal protein L9: evidence for rapid two-state folding.** *Biochemistry* 1998, 37:1025-1032.

79. De Laureto PP, Scaramella E, De Filippis V, Marin O, Doni MG, Fontana A: **Chemical synthesis and structural characterization of the RGD-protein decorsin: a potent inhibitor of platelet aggregation.** *Protein Sci* 1998, 7:433-444.

80. Johansson JS, Gibney BR, Skalicky JJ, Wand AJ, Dutton PL: **A native-like three-helix bundle protein from structure-based redesign: a novel maquette scaffold.** *J Am Chem Soc* 1998, 120:3881-3886.

81. Canne LE, Simon RJ, Kent SBH: **Solid phase protein synthesis by chemical ligation of unprotected peptides in aqueous solution.** In *Proceedings of the American Peptide Symposium, 1997; Nashville.* Edited by Tam JP. ESCOM; 1998:in press.

82. Berman AL, Kolker E, Trifonov EN: **Underlying order in protein sequence organization.** *Proc Natl Acad Sci USA* 1994, 91:4044-4047.

83 •• Muir TW, Sondhi D, Cole PA: **Expressed protein ligation: a general method for protein engineering.** *Proc Natl Acad Sci USA* 1998, in press.
This paper describes the preparation of proteins in which part of the protein molecule comes from recombinant DNA expression, and part from chemical peptide synthesis. Sure to be popular with the cloning crowd as a way of overcoming some of the limitations of recombinant DNA expression.

84. Severinov K, Muir TW: **Expressed protein ligation: a novel method for studying protein-protein interactions in transcription.** *J Biol Chem* 1998, in press.

85. Erlanson DA, Chytil M, Verdine GL: **The leucine zipper domain controls the orientation of AP-1 in the NFAT•AP-1•DNA complex.** *Chem Biol* 1996, 3:981-991.

86. Muir TW, Dawson PE, Fitzgerald MC, Kent SBH: **Probing the chemical basis of binding activity in an SH3 domain by protein signature analysis.** *Chem Biol* 1996, 3:817-825.

87 • Dawson PE, Fitzgerald MC, Muir TW, Kent SBH: **Methods for the chemical synthesis and readout of self-encoded arrays of polypeptide analogues.** *J Am Chem Soc* 1997, 119:7917-7927.
This paper contains a compendium of chemical protein methods: novel array synthesis; a novel method for self-encoded libraries; functional selection; and one-step mass spectrometric read out/decoding. The basis for a complete protein chemistry in one publication.

88. Wallin E, von Heijne G: **Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms.** *Protein Sci* 1998, 7:1029-1038.

89. Muir TW, Williams MJ, Ginsberg MH, Kent SBH: **Design and chemical synthesis of a neoprotein structural model for the cytoplasmic domain of a multisubunit cell-surface receptor: integrin allbb3 (platelet GPIIb-IIIa).** *Biochemistry* 1994, 33:7701-7708.

Annu. Rev. Biochem. 2000. 69:923–60

SYNTHESIS OF NATIVE PROTEINS BY CHEMICAL LIGATION*

Philip E. Dawson[1] and Stephen B. H. Kent[2]

[1]*The Scripps Research Institute, La Jolla, California 92037; e-mail: dawson@scripps.edu;*
[2]*Gryphon Sciences, South San Francisco, California 94080; e-mail: skent@gryphonsci.com*

Key Words chemical protein synthesis, thioester, protein, peptide, solid phase synthesis, polymer-supported synthesis, protein engineering

■ **Abstract** In just a few short years, the chemical ligation of unprotected peptide segments in aqueous solution has established itself as the most practical method for the total synthesis of native proteins. A wide range of proteins has been prepared. These synthetic molecules have led to the elucidation of gene function, to the discovery of novel biology, and to the determination of new three-dimensional protein structures by both NMR and X-ray crystallography. The facile access to novel analogs provided by chemical protein synthesis has led to original insights into the molecular basis of protein function in a number of systems. Chemical protein synthesis has also enabled the systematic development of proteins with enhanced potency and specificity as candidate therapeutic agents.

CONTENTS

*At the time of the invitation, as now, Stephen Kent is President and Chief Scientist at Gryphon Sciences. Gryphon Sciences is focused on the development and sale of enhanced protein therapeutics using chemical protein synthesis. The core technology of the company is largely the subject matter of the chapter we have submitted.

INTRODUCTION: Protein Science in the Postgenome Era

An important current objective in biomedical research is to understand the molecular basis of the numerous and intricate biological activities of proteins and therefore to be able to predict and control these activities. The importance of this goal is dramatically increased today because of the explosive success of the genome-sequencing projects, which have revealed hundreds of thousands of new proteins, but only as predicted sequence data (1). For the biologist, elucidation of the biological function of a predicted protein molecule is thus a challenge of great significance. In the words of Freeman Dyson, "[In the post-genome era], proteins will emerge as the big problem and the big opportunity. When this revolution occurs, it will have a more profound effect than the Human Genome Project on the future of science and medicine" (2).

For the past 20 years, most studies of the molecular basis of protein action have been carried out by recombinant DNA-based expression of proteins in genetically engineered cells (3). From its introduction, this powerful method revolutionized the study of proteins by enabling the production of large amounts of proteins of defined molecular composition and by allowing the systematic variation of the amino acid sequence of proteins (4). Expression of proteins in engineered cells is now a mature technology, and its scope and limitations are well understood: (*a*) Small proteins (i.e. <30 kDa) are easier to express than large, multidomain proteins; (*b*) folding of large-protein molecules can also be a challenge; (*c*) product

heterogeneity is frequently a problem, caused by uncontrolled processing of the nascent polypeptide in the cell; and (*d*) the overexpression of proteins that are toxic to the cell, such as proteases, can be difficult (5).

Additionally, because the cell is used as a protein factory, such molecular biology studies are inherently limited to the 20 genetically encoded amino acids. Efforts have been made to use cell-free synthesis to expand the repertoire of ribosomal synthesis to include noncoded amino acids as building blocks (6, 7). These attempts to incorporate other amino acids have had very limited success—obtaining adequate amounts of pure protein from the cell-free translation systems can be a significant challenge (8), and many unnatural amino acids are simply not compatible with ribosomal polypeptide synthesis (9).

Chemical synthesis is an attractive alternative to biological methods of protein production. The use of synthetic chemistry promises the unlimited variation of the covalent structure of a polypeptide chain with the objective of understanding the molecular basis of protein function. Chemistry also promises the ability to systematically tune the properties of a protein molecule in a completely general fashion.

This vision was one of the prime imperatives of organic chemistry in the time of Emil Fischer at the beginning of the 20th century. In a 1905 letter to Adolf Baeyer, Fischer wrote, "My entire yearning is directed toward the first synthetic enzyme. If its preparation falls into my lap with the synthesis of a natural protein material, I will consider my mission fulfilled" (10). In the decades since then, the challenge of applying the methods of chemistry to the study of protein action has stimulated numerous advances in synthetic methods. Historically, these advances included the use of novel reversible protecting groups (11), novel activation methods for the formation of covalent bonds (12), and even polymer-supported synthesis (13), all of which sprang from the drive to apply the science of chemistry to the study of proteins.

DOMAINS: Building Blocks of the Protein World

Because proteins are large molecules, applying chemical synthesis to them is a considerable challenge. Furthermore, the biological functions of proteins originate in the tertiary structure of the protein molecule—that is, in the precise three-dimensional folded structure of the polypeptide chain. The typical protein molecule is ~30 kDa in size and consists of two ~15-kDa domains (14–16); each domain has a polypeptide chain length of ~130 (±40) amino acids (14–16). Protein domains are defined as autonomous units of folding and, frequently, of function (17, 18). As such, domains are the building blocks of the protein world. The challenge confronting the chemist is, first, the total synthesis of folded domains and then the ability to stitch these domains together to build complex protein molecules.

CHEMICAL PROTEIN SYNTHESIS: The State of the Art in 1990

Since last reviewed in this journal (19), total chemical synthesis of native proteins has made a number of important contributions to biomedical research. It is notable that the Kent laboratory at the California Institute of Technology used total chemical synthesis based on predicted gene sequence data to carry out pioneering studies of human immunodeficiency virus 1 (HIV-1) protease enzyme (20). The existence of this virally encoded aspartyl proteinase had been postulated based on an analysis of viral nucleic acid sequence data, and molecular genetic studies had indicated that its action in processing the gag-pol polyprotein was essential to the viral life cycle (21). For this reason, the HIV-1 protease was, early on, proposed as an important target for drug development. The first preparations of the enzyme of defined molecular composition were produced by chemical synthesis (22), using a highly optimized version of stepwise solid-phase peptide synthesis (19). This work proved that the active form of the HIV-1 protease was a homodimer consisting of two identical 99-residue polypeptide chains, and it showed that the chemically synthesized enzyme accurately processed the putative cleavage sites in the viral gag-pol translation product (22).

In a strikingly important contribution, total chemical synthesis was also used to prepare large amounts of homogeneous enzyme for the determination of the original crystal structures of the HIV-1 protease molecule (Figure 1). The structure of the unliganded synthetic enzyme (23) corrected a seriously flawed low-resolution structure (24) that had been obtained by using protein derived from recombinant expression in *Escherichia coli*. Even more significantly, use of chemically synthesized enzyme provided the first high-resolution cocrystal structures of the HIV-1 protease molecule complexed with substrate-derived inhibitors (25–27). These structural data were made freely available to the research community and formed the foundation for the successful worldwide programs of structure-based drug design (28) that led to the development of the highly effective protease inhibitor class of acquired immune deficiency syndrome therapeutic agents (29).

SYNTHETIC-PEPTIDE CHEMISTRY: Useful but Bounded

Despite successful syntheses of the HIV-1 protease (22) and of a limited number of other proteins (30–35), at the start of the decade of the 1990s total chemical synthesis, by the standard methods of peptide chemistry of even a small protein molecule remained a daunting task, often requiring large teams and taking years to complete, with no guarantee of success. The routine, reproducible preparation of synthetic polypeptides of defined chemical structure was limited to products of ~50 amino acid residues (19; Figure 2). This size limitation applied equally to synthesis by solution or by solid-phase methods, but for differing reasons.

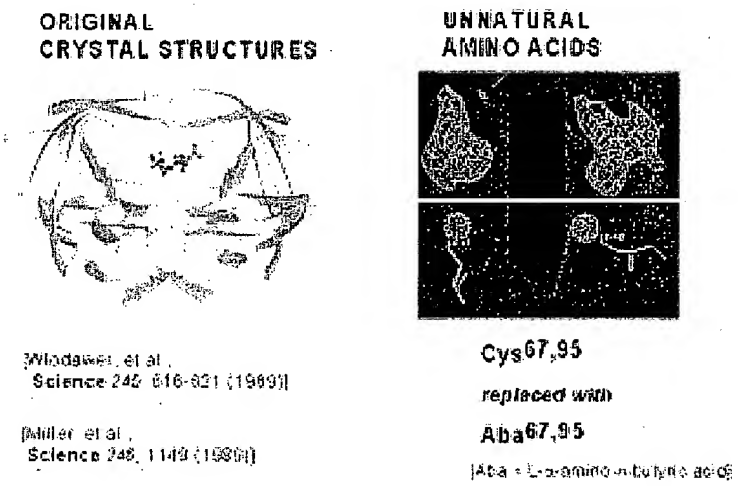


Figure 1 Crystal structures of chemically synthesized HIV-1 protease. These were the original high-resolution structures (23, 25–27) of this protein and guided the subsequent drug design programs. The synthetic protein preparation used for X-ray crystallography contained L-α-amino-*n*-butyric acid residues in place of the two Cys residues in each subunit. (*Left*) Molscript representation of the synthetic enzyme in complex with the substrate-derived inhibitor MVT101 (25). (*Right upper panel*) 2Fo-Fc electron density map for the side chains of the unnatural amino acids used to replace the two Cys residues in each subunit of the synthetic enzyme (23). (*Lower panel*) Side chains of the L-α-amino-*n*-butyric acid residues superimposed on the mercury atoms from Cys-containing enzyme (24) that has been crystallized in the same space group. This shows that the side chains of the unnatural amino acid have the same conformation as the natural Cys side chains. (Adapted from References 23 and 25).

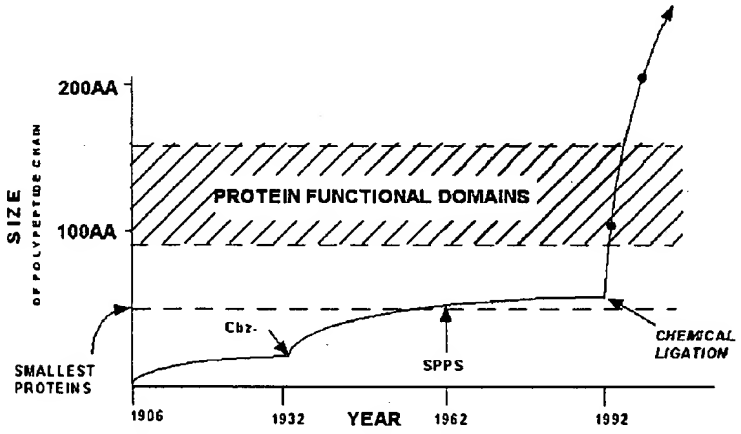


Figure 2 Historical progress in the size of synthetically accessible polypeptides.

Classical solution synthetic chemistry involves the preparation of fully protected peptide segments and their subsequent condensation in organic solvents for the convergent synthesis of large polypeptides (36). The problems associated with this classical approach have been summarized (19). These limitations include the laborious and technically demanding preparation of the protected segments, the lack of general, highly resolving methods for the purification of protected segments, and the inability to directly characterize fully protected peptides—even by modern analytical methods.[1] In addition, it became apparent that fully protected polypeptide chains frequently had only limited solubility in organic solvents that are useful for peptide synthesis. This poor solubility made such protected peptide segments difficult to work with, and the low concentrations attainable for reacting segments often led to slow and incomplete reactions (37, 38).

By contrast, unprotected peptide segments usually have good solubility properties, are more easily handled, and can be directly characterized. The most efficient way of making unprotected peptides is stepwise solid-phase peptide synthesis (SPPS). This ingenious chemical synthesis method, the progenitor of all polymer-supported organic chemistry, was introduced in 1963 by Merrifield (13). Both the principles and the practical aspects of SPPS have been thoroughly described (19). By the end of the 1980s, it was possible by highly optimized stepwise SPPS (19) to make, in good yield and high purity, essentially any peptide $\leq$50 amino acids in length. Reverse-phase high-pressure liquid chromatography methods could be routinely used to purify these synthetic products and to evaluate their homogeneity (39). More recently, electrospray mass spectrometry has provided a straightforward general method for the precise characterization of the covalent structure of unprotected synthetic peptides (40). Despite the extraordinary power of solid-phase peptide synthesis, lack of quantitative reaction eventually leads to the formation of significant levels of resin-bound byproducts. It is this statistical accumulation of coproducts that limits the ultimate size of high-purity polypeptides of defined covalent structure that can be effectively prepared in this way.

Thus, synthetic peptide chemistry, whether by stepwise SPPS (19) or by solution methods (36), can provide routine access to polypeptide chains of ~50 amino acids. This corresponds to only the very smallest proteins and protein domains.

A number of attempts were made to take advantage of the ability to make, characterize, and handle unprotected peptides (41–44). Noteworthy is the development of enzymatic ligation methods for the preparation of large polypeptides from synthetic peptide segments, with ligase enzymes specifically engineered for this purpose by the methods of molecular biology (45). Ironically, the principal obstacle to general utility of enzymatic ligation has proven to be the limited

[1]For example, electrospray mass spectrometry has become one of the most useful tools for determining the covalent structure of peptides (40); this powerful method involves direct ionization of an analyte from aqueous solution. The efficacy of this ionizaton depends on the presence of multiple ionizable groups in the molecule under study. The lack of such groups in fully protected peptides precludes direct analysis by electrospray mass spectrometry.

solubility of even the unprotected peptide segments, under the physiological conditions compatible with the enzymes used (46). Despite considerable efforts and some notable successes (47), such methods have not found widespread use.

CHEMICAL LIGATION OF UNPROTECTED PEPTIDE SEGMENTS

As recently as 1991 (48), the challenge remained: namely, to develop methods that enable the general application of the tools of chemistry to the world of the protein molecule. It was evident (41–44, 48) that a truly useful approach to chemical protein synthesis would be based on the ability to routinely make unprotected peptides ≤50 amino acid residues in length and would consist of a practical way to stitch such synthetic peptides together to give polypeptides of any desired length, and hence the corresponding folded protein molecules.

Based on this premise, in the early 1990s the principle of chemoselective reaction (49) was adapted to enable the use of unprotected peptide segments in chemical protein synthesis (50). This novel "chemical ligation" approach relied on a conceptual breakthrough, the principles of which are shown in Figure 3.

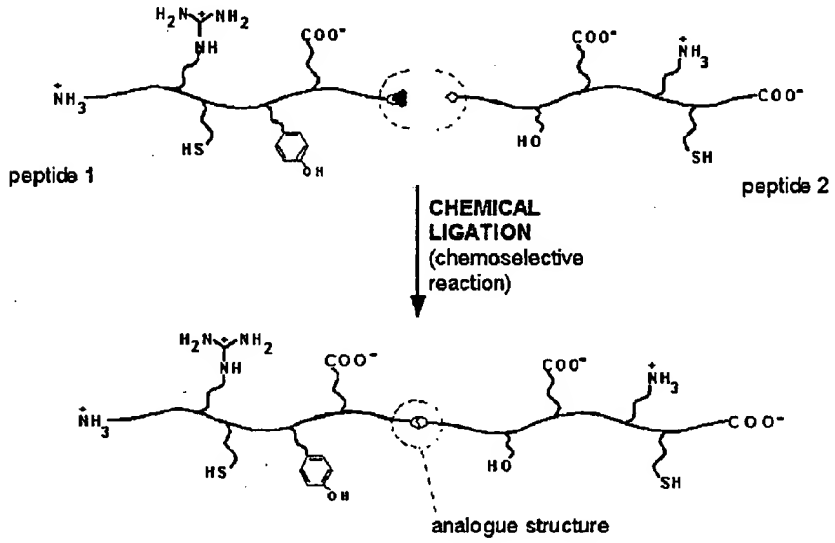


Figure 3 Principles of chemical ligation (48, 50). Uniquely reactive functionalities are incorporated into each peptide by chemical synthesis. Mutual chemoselective reaction of these moieties allows the use of completely unprotected peptide segments, which are prepared by standard means and can be readily purified and characterized by sensitive, high-resolution methods. Reaction is carried out in aqueous solution in the presence of chaotropes, such as 6 M guanidine-HCl, to increase the concentration of reacting segments and speed up the reaction. The product polypeptide is obtained directly in final form.

TABLE 1 Chemistries used for the synthesis of native proteins by chemical ligation of unprotected peptide segments

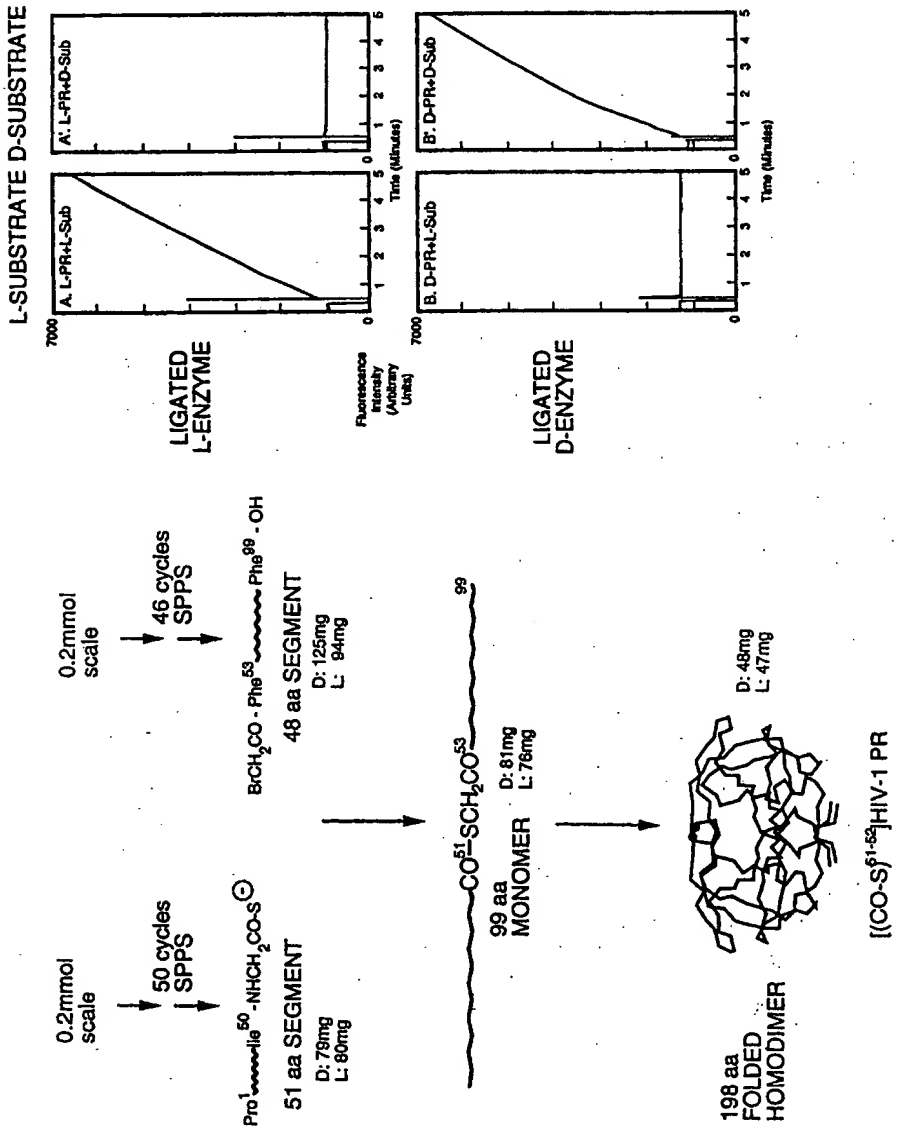
Chemistry	Reference
1. Thioester-forming ligation	50
2. Oxime-forming ligation	53
3. Thioether-forming ligation	59
4. Directed disulfide formation	85
5. Thiazolidine-forming ligation	60, 61
6. Peptide bond-forming ligation	62

In essence, the use of unique, mutually reactive functional groups not normally found in peptides enabled the site-specific ligation of completely unprotected peptide segments for the synthesis of large polypeptide chains. Reactions were designed to be carried out in aqueous solution, and a chaotropic agent such as 6 M guanidine-HCl was used to increase the solubility of the reacting peptide segments, thereby allowing the use of higher peptide concentrations to accelerate the ligation reactions.

This chemical ligation method has proven to be simple to implement, highly effective, and generally applicable (51). A variety of ligation chemistries has been used (Table 1), and the chemical ligation of unprotected peptide segments has provided access to a range of protein targets.

The price paid for such unprecedented synthetic convenience, at least in the initial stages of development of the method, was the formation of an unnatural structure at the site of ligation between two peptide segments (50). However, these unnatural structures are often well-tolerated within the context of a folded protein, and numerous examples exist of fully active protein molecules that are chemically synthesized in this way. Some early examples of proteins prepared by the chemical ligation method include (*a*) enzymatically active HIV-1 protease (50); (*b*) the mirror image enzyme D-HIV-1 protease, which was prepared by a thioester-forming chemical ligation (52; Figure 4) and its high-resolution crystal structure determined (20); (*c*) the facile total synthesis of proteinlike TASP molecules of unusual topology (53–55); (*d*) the synthesis of backbone-engineered variants of the HIV-1 protease (56) to investigate the mechanism of the enzyme (Figure 5); (*e*)

Figure 4 Total synthesis of mirror image forms of the HIV-1 protease enzyme molecule (52). (*Left*) Unprotected ~50-residue peptide segments are reacted by thioester-forming chemical ligation to give the 99-residue polypeptide chain of the HIV-1 protease monomer. Folding gave excellent yields of the homodimeric enzyme molecules. (*Right*) Reciprocal chiral specificity of the mirror image enzyme molecules, exemplified in a fluorogenic assay. The ligated L-enzyme acted only on the L-substrate, whereas the ligated D-enzyme acted only on the D-substrate.



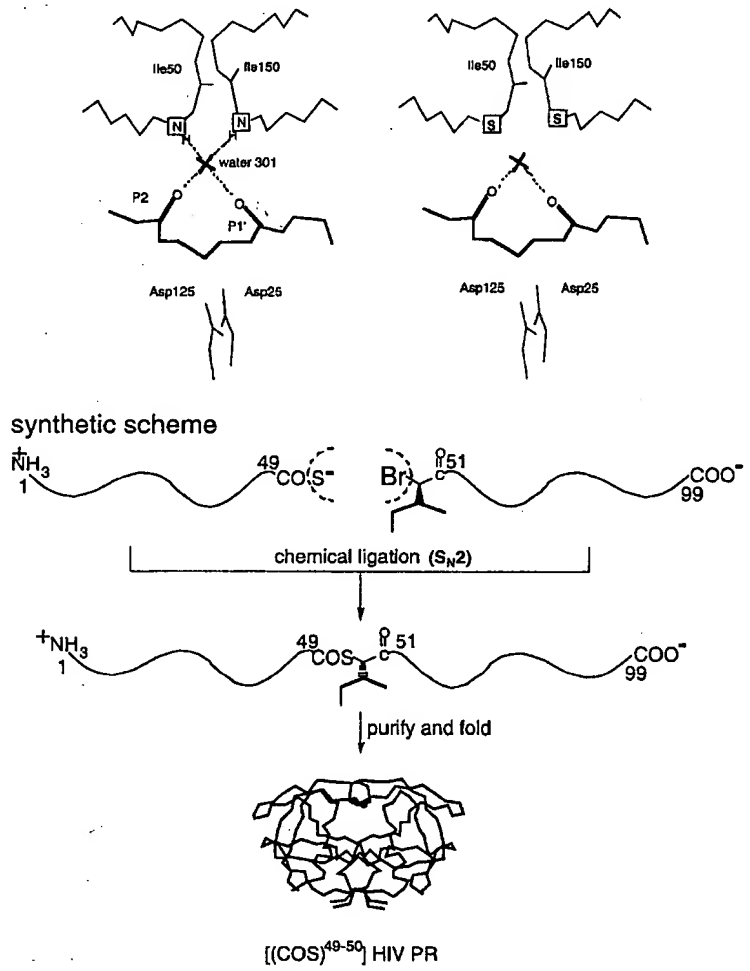


Figure 5 Backbone-engineered HIV-1 protease by chemical ligation (56). (*Top*) Design of the variant enzyme. (*Top left*) H bonding of "water 301" by the amide —NH— of Ile50 at the tip of each flap structure. (*Top right*) Sulfur atoms replacing these —NH— moieties, thus deleting the H-bonding potential. (*Bottom*) Synthetic scheme. Nucleophilic thioester-forming ligation, with inversion of configuration at the 'D-Ile50' chiral center, to give the desired 99-residue polypeptide, which is folded to form the homodimeric enzyme molecule.

the synthesis of fully functional covalent heterodimers of b/HLH/Z transcription factors (57; Figure 6); and (*f*) the synthesis of receptor mimetics (58).

These and other syntheses performed by the chemical ligation method demonstrated that proteins could now be made in high yield and good purity from unprotected peptide building blocks and that unnatural analogs could be readily prepared to investigate new aspects of protein structure and function.

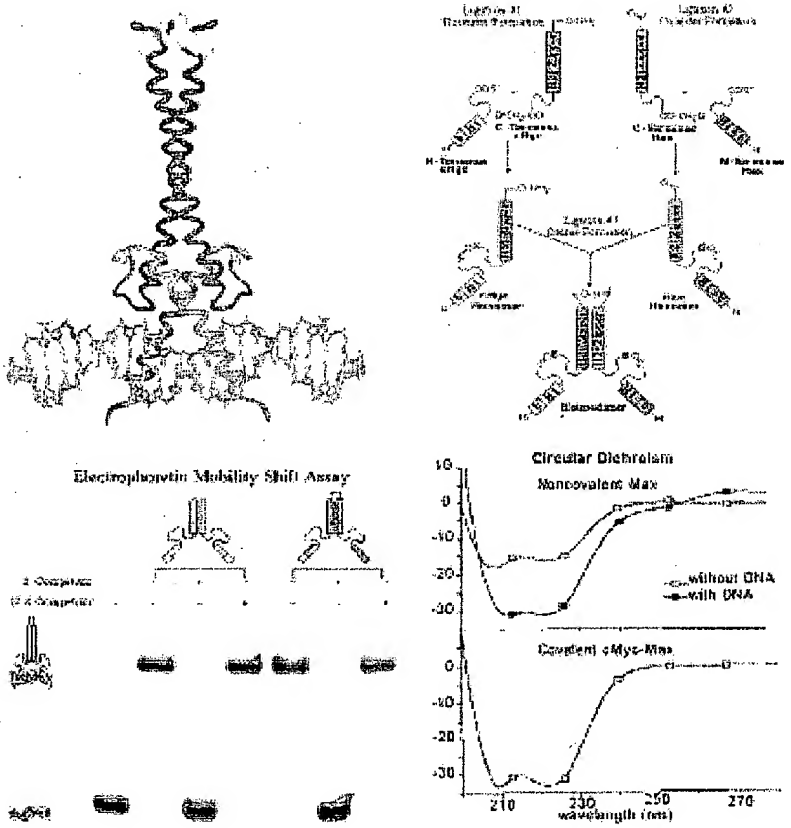


Figure 6 Total synthesis of a covalent heterodimeric transcription factor, cMyc-Max, by convergent chemical ligation (57). (*Upper left*) Molecular model of the covalent construct, bound to cognate duplex DNA. (*Upper right*) Synthetic scheme—each B/HLH/Z domain was assembled by thioester-forming chemical ligation of two peptide segments; these polypeptide products were then covalently linked by oxime-forming chemical ligation to yield a synthetic protein construct with two N terminals and no C terminal. (*Lower right*) Circular dichorism measurements showed that the covalent cMyc-Max construct folded correctly and was preordered even in the absence of cognate DNA. (*Lower left*) The covalent cMyc-Max heterodimer was active in a gel shift assay for DNA binding. Adapted from Reference 57 and Ferré-D'Amare AR (1995. PhD thesis).

NATIVE CHEMICAL LIGATION

The original ligation chemistries (50, 53, 59–61) gave a nonpeptide bond at the site of ligation. In 1994, based on the original principles of the chemical ligation method (48, 50), Dawson et al introduced an ingenious extension of the chemistries used for the chemoselective reaction of unprotected peptide segments—native chemical

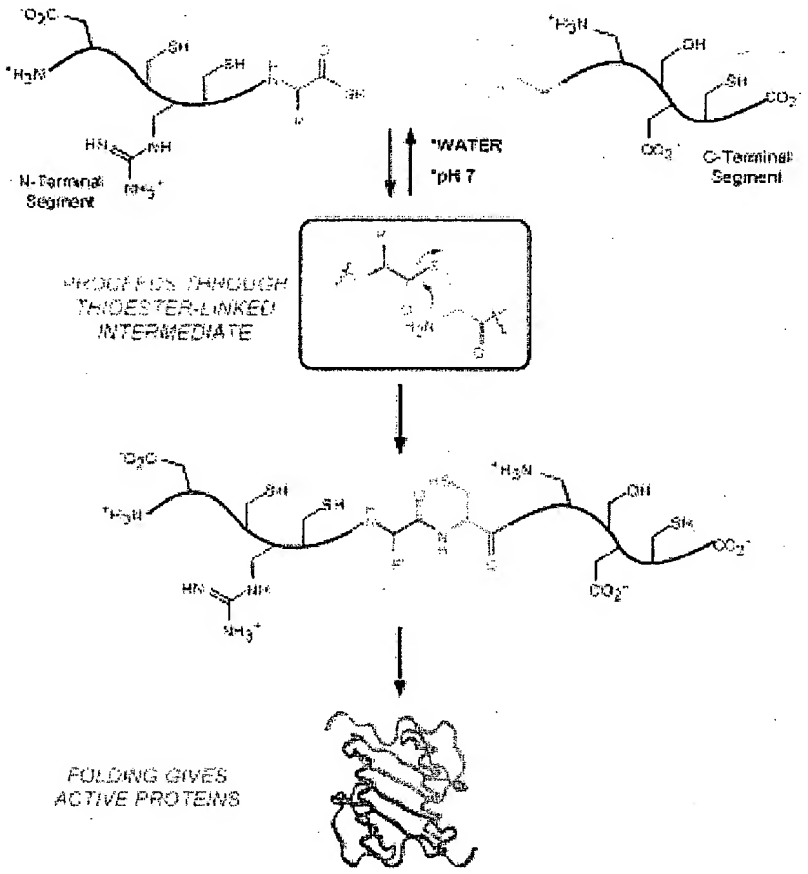


Figure 7 Native chemical ligation (62). Unprotected peptide segments are reacted by means of reversible thiol/thioester exchange to give thioester-linked initial reaction products. Uniquely, the thioester-linked intermediate involving an N-terminal Cys residue (*boxed*) is able to undergo nucleophilic rearrangement by a highly favored intramolecular mechanism; this step is irreversible (under the conditions used) and gives a polypeptide product that is linked by a native amide (i.e. peptide) bond. Only a single reaction product is obtained, even in the presence of additional Cys residues in either segment. The product polypeptide is subsequently folded to give the desired synthetic protein molecule.

ligation (62). In this method, simply mixing together two peptide segments that contain correctly designed, mutually reactive functionalities led to the formation of a single polypeptide product containing a native peptide bond at the ligation site. This highly chemoselective reaction is performed in aqueous solution at neutral pH under denaturing conditions. The chemical principles underlying the native chemical ligation method are shown in Figure 7.

The essential feature of native chemical ligation is the (transient) formation of a thioester-linked product, as was the case in the original method (50) for the

synthesis of proteins by chemical ligation. In the native chemical ligation method, however, this initial thioester-linked product is not isolated; rather, it is expressly designed to undergo spontaneous rearrangement, via intramolecular nucleophilic attack, to give the desired amide-linked product (62, 63). The result is a completely native polypeptide chain that is obtained directly in final form.

A feature of the native chemical ligation method is that ligation occurs at a unique N-terminal Cys residue. It does not matter how many additional internal Cys residues are present in either segment (62, 64). No protecting groups are necessary for any of the side-chain functional groups normally found in proteins, and quantitative yields of the ligation product are obtained.

Where this exquisite selectivity originates is important; it lies in the use of reversible thiol/thioester exchange reactions to form the thioester-linked intermediate ligation products (62, 63). The exchange is promoted by suitable thiol catalysts and is freely reversible under the neutral aqueous conditions used for the reaction. Intramolecular nucleophilic attack to form the amide bond at the ligation site is irreversible under the same conditions, so that, over the time course of the reaction, all of the freely equilibrating intermediates are depleted by the irreversible reaction step, giving a single polypeptide ligation product. Typical data from a native chemical ligation reaction are shown in Figure 8. Detailed studies of mechanistic aspects of the native chemical ligation reaction have been published (63, 65).

Formation of a native peptide bond at the ligation site has been unequivocally demonstrated by a variety of methods, including chemistry (62), NMR (66), and X-ray crystallography (67; Figure 9). A remarkable feature of the native chemical ligation of unprotected peptide segments is the absence of racemization in the coupling reaction. Detailed studies have been carried out, and no racemization was detected in the ligation product to a limit of $<1\%$ D-amino acid content (68).

BIOCHEMICAL PEPTIDE LIGATION

Protein Splicing

This cellular processing event occurs post-translationally at the polypeptide level in certain classes of protein molecules, to generate a truncated final product that results from excision of the central portion of the initial polypeptide produced on the ribosome. Intein-mediated protein splicing is a biochemical[2] reaction that

[2]It can be expected that examples of chemical ligation of protein domains will be discovered in vivo, making use of biochemical mechanisms other than intein-mediated protein splicing. In a variety of phyla, the cell already makes use of polypeptide thioesters in numerous biochemical processes. These processes include the action of cysteine proteinases (69); the ubiquitination of proteins targeted for catabolic destruction (70); the nonribosomal synthesis of peptides (71); and in the complement-mediated response to foreign pathogens (72). Given the obvious utility to the cell of cutting and pasting protein domains at the polypeptide level, it is reasonable to assume that nature will have worked out ways of taking advantage of its existing "tool kit" to accomplish this task by chemical ligation.

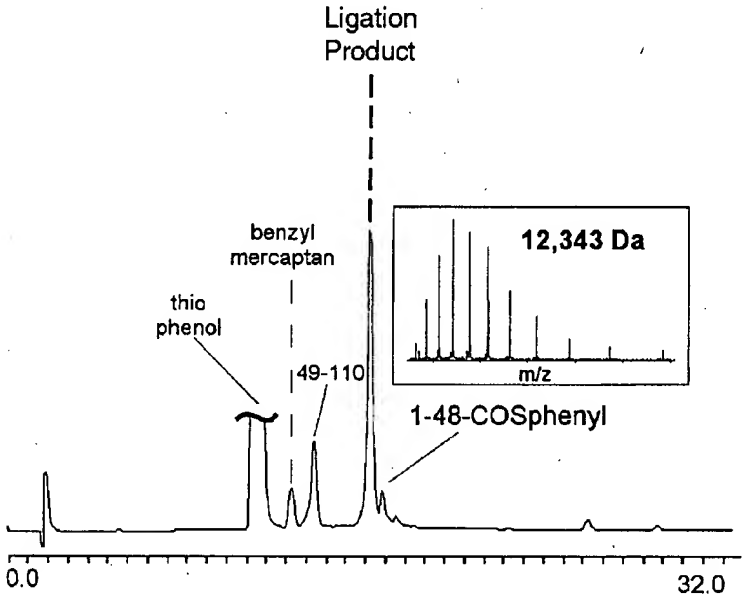
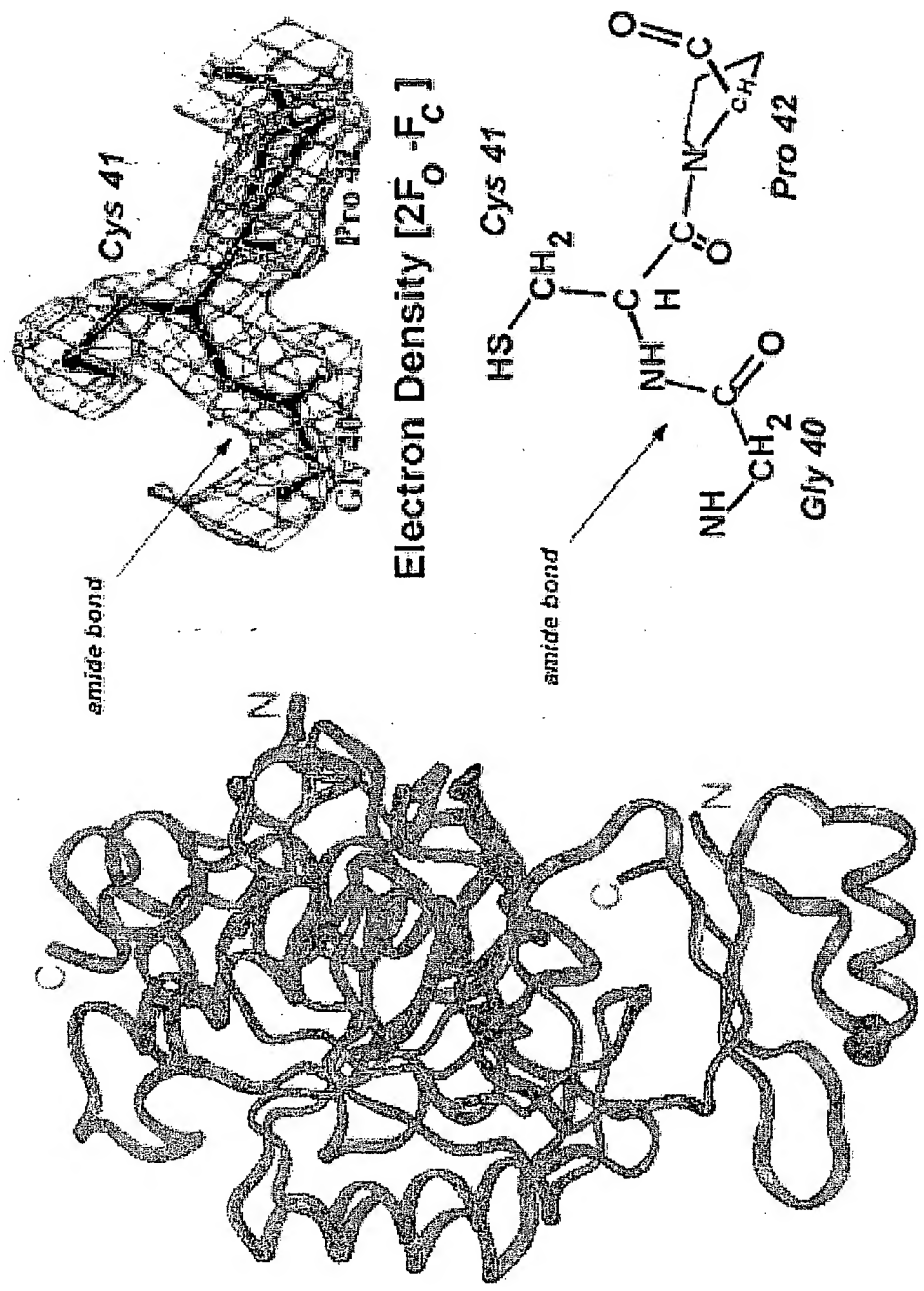


Figure 8 Raw analytical HPLC data from the synthesis by native chemical ligation of the 110-residue polypeptide chain of the enzyme barnase (63). The two unprotected peptide segments, (1-48)$^{\alpha}$COSbenzyl and Cys49-110, were reacted in aqueous solution at pH 7 in the presence of a thiol catalyst. Data after 7 h show a nearly quantitative reaction to form a single product. (*Insert*) Electrospray mass spectrometric data for the ligated 110-residue polypeptide, M_w 12,343.

has considerable utility in its own right, and it is the subject of another chapter in this same volume (73). It is interesting that the publication of the native chemical ligation method (62) in 1994 strongly influenced the subsequent elucidation of critical aspects of the biochemical mechanism of natural protein splicing, as described by Xu et al (74). Based in part on an appreciation of the thioester-mediated acyl shift mechanism that had already been defined for the synthetic native chemical ligation reaction (62), protein splicing was shown to proceed via (intein-mediated) formation of (thio)ester-linked intermediates, followed by nucleophilic attack to form the final amide-linked spliced polypeptide (73).

In fact, the mechanisms of native chemical ligation and intein-mediated protein splicing are quite distinct in certain critical respects, despite the shared features

Figure 9 Crystal structure of the synthetic protein Eglin C. (*Left*) Tserine protease inhibitor Eglin C (*orange*) complexed with recombinant subtilisin (*green*) (67). (*Right*) Structure of the bond formed at the site of native chemical ligation in the synthesis of the Eglin C protein. The newly formed amide bond is defined by continuous electron density in the $2F_o$-F_c map. Adapted from Reference 67.



of involvement of (thio)ester-linked intermediates and the final step of an *S*- or *O*- to *N*-acyl shift to form the amide-linked product. Both the mechanistic basis of selectivity and the thermodynamic driving force for the ligation reaction differ between the two processes. In intein-mediated splicing, the precise site of joining the N-extein and C-extein peptides is biochemical in origin and arises from spatial juxtaposition of the reacting residues, which is brought about by the folded conformation of the intein protein domain. By contrast, in native chemical ligation, initial reaction products may involve every thiol functionality in the reacting peptide segments (64); the exquisite selectivity originates in freely reversible thiol/thioester exchange among these initial products, followed by irreversible rearrangement of just one intermediate to give a single, defined reaction product (62).

The thermodynamic driving force also has distinct origins in the two processes. The starting point for intein-mediated splicing is amide bonds within a single polypeptide chain, and the reaction yields a ligated product and a peptide fragment (73). Thus, the driving force for the biochemical splicing reaction is actually the same as for solvolytic cleavage of peptide bonds—the generation from an uncharged amide of an ionized moiety (perhaps two) with favorable solvation properties. There may also be a contribution from the distorted (high energy) state of the starting amide bond induced by the folded structure of the intein-containing protein (73). For native chemical ligation, the reaction of a peptide-thioester with an amine to form an amide (i.e. peptide) bond is strongly favored on simple enthalpic grounds. Because of these critical mechanistic differences in thermodynamic driving force and selectivity of reaction, it is an oversimplification to describe native chemical ligation as "biomimetic" (75).

Use of defective-intein expression systems as a route to the preparation of peptide-thioesters for use in native chemical ligation is discussed on p. 951.

Conformationally Assisted Ligation

In some cases, folding conditions can be used to accelerate the rate of native chemical ligation (76). Many proteins can be cut into two or more polypeptides that can be reconstituted to form a nativelike conformation. In these cases, the revealed N and C terminals of the peptide fragments are located in close proximity at the site of chain scission. This greatly increases the collision frequency, and a weakly activated C-terminal group such as a thioester can be used to religate the fragments. Total synthesis of proteins using this approach has been demonstrated with the chymotrypsin inhibitor CI2 (Figure 10). When two synthetic peptide segments spanning the CI2 molecule, one incorporating a C-terminal thioester and the other an N-terminal cysteine, are mixed together under folding conditions, conformationally-assisted ligation proceeds in <2 min, compared with several hours for chemical ligation under denaturing conditions (76). In suitable systems, the peptide-$^{\alpha}$thioester segment can even be mixed under folding conditions with a version of the other peptide segment without a Cys at the N terminal, and conformationally-assisted ligation still proceeds in a matter of hours.

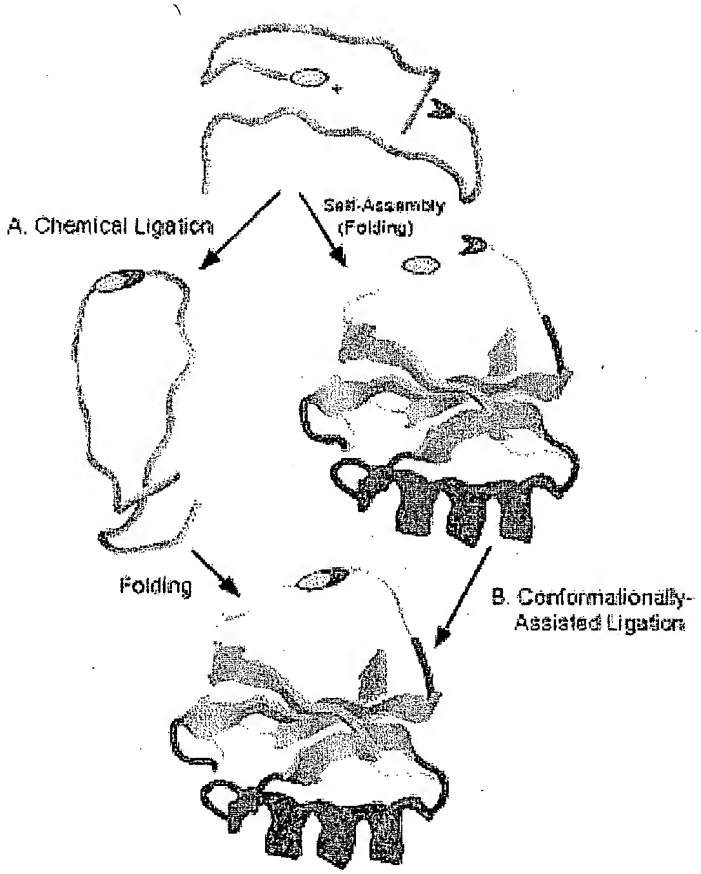


Figure 10 Conformationally assisted chemical ligation, exemplified for the chymotrypsin inhibitor C12. (*A, left*) Thioester-mediated chemical ligation at Cys under standard denaturing conditions occurs over several hours. (*B, right*) The same ligation reaction under folding conditions, in which the two segments associate to increase the collision frequency between the reacting functionalities, is complete within 3 min (76).

This conformationally assisted chemical ligation extends previously developed semisynthetic approaches that used other weakly activated ester groups for C-terminal activation of a peptide segment (77).

SCOPE OF NATIVE CHEMICAL LIGATION FOR THE SYNTHESIS OF PROTEINS

The broad scope of the native chemical ligation method for the total synthesis of proteins is summarized in Figure 11 and by the data shown in Table 2.

The first applications of native chemical ligation were to small, Cys-rich proteins such as disulfide-cross-linked secretory proteins or the zinc-finger proteins.

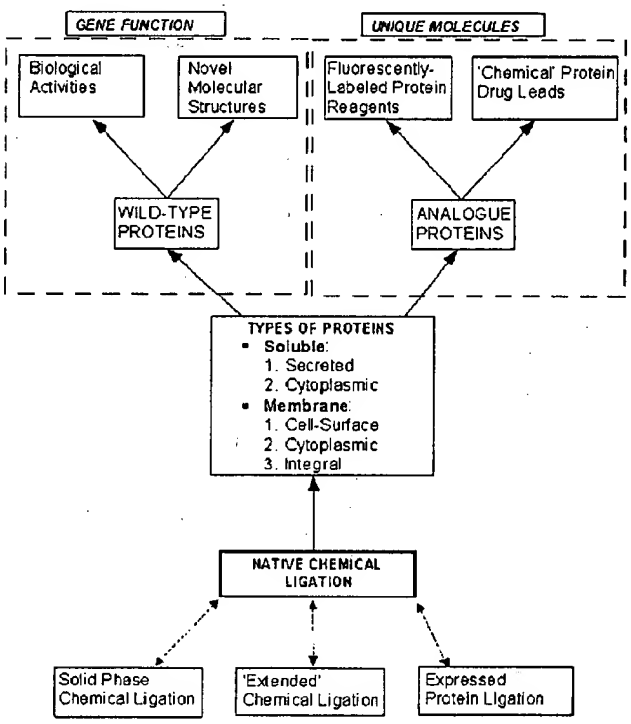


Figure 11 Applications of native chemical ligation.

In all, >300 biologically active proteins from >20 different families have been successfully prepared by total chemical synthesis with this method. These are still early results in what will surely be more widespread application of the method, but they demonstrate routine synthetic access to single-domain proteins and suggest that native chemical ligation will provide the basis of a general synthetic access to the world of proteins.

FOLDING SYNTHETIC PROTEINS

The activity of a protein molecule originates in the precise tertiary structure of the folded polypeptide chain. To complete the total synthesis of a functional protein molecule, the synthetic polypeptide chain that corresponds to the sequence of the protein must be folded to form the correct three-dimensional structure. Our intriguing experience to date has been that chemically synthesized polypeptide

TABLE 2 Selected proteins prepared by total synthesis using native chemical ligation[a]

Protein class/families[b]	Protein molecular mass (kDa)	Polypeptide size (aa)
Secretory		
Chemokines	8–10	~70
Cytokines	15–20	~160
BMPs	~25	2 × 115
Ser PR inhibitors	6–8	58–70
Agouti proteins	6–12	50–112
AFP	~6	~50
Anaphlyatoxins	~8	~70
EGFs/TGF-α	~6	~50
Receptor/membrane		
β_2 microglobulin	12	99
glp1r N-term domain	14	120
Influenza m2	50	4 × 97
Intracellular		
SH2 domains	~10	~90+
SH3 domains	~7	~60
b/HLH/Z	16–20	2 × 70–180
Zn-finger	~8	~70
Redox		
Desulforedoxin	8	2 × 36
Rubredoxin	6	53
Cyt b5	10	82
Enzymes		
Retroviral proteases	20	2 × 99–116
Secretory PLA2s	14	~120
MIF	39	3 × 115
Barnase	12	110

[a]Research scale synthesis typically gives 50–100 mg of each protein; each of the above proteins had the expected biochemical or biological activity; three-dimensional molecular structures were determined by NMR or X-ray crystallography for many of these proteins.

[b]BMP, bone morphogenetic protein; AFP, anti-fungal protein; EGF, epidermal growth factor; TGF-α, transforming growth factor-α; MIF, macrophage migration inhibitory factor.

TABLE 3 Some structural motifs successfully folded as synthetic proteins[a]

Chemokine fold	SH3
Ser protease inhibitor fold	PLA2 (14-kDa form)
Kringle fold	Cytokine fold
Agouti Cys-rich domain	TGF-β fold
EGF-fold	b/HLH/Z DNA-binding domains
SH2	Rubredoxin
Zn-fingers	4OT/MIF
Aspartyl protease fold	Chitin binding domains
Anaphylatoxins	Ion channels

[a]EGF, Epidermal growth factor; TGF-β, transforming growth factor β; 4OT/MIF, 4-oxalocrotonate tantomerase/macrophage migration inhibitory factor.

chains fold efficiently in vitro to give fully functional protein molecules (78). Such synthetic proteins have unique, defined folds of the polypeptide chain, as shown by NMR measurements (79, 80) and by X-ray crystallography (67, 79, 81; Figures 12 and 13). Some examples of structural motifs successfully folded as synthetic proteins are given in Table 3.

Thus, correct folding of synthetic proteins is efficient, accurate, and general at the level of single domains.

A growing list of multidomain proteins have also been successfully produced by the folding of chemically synthesized polypeptide chains (e.g. see Figures 1, 6, and 14). These proteins include homodimers (22, 23), heterodimers (20, 57),

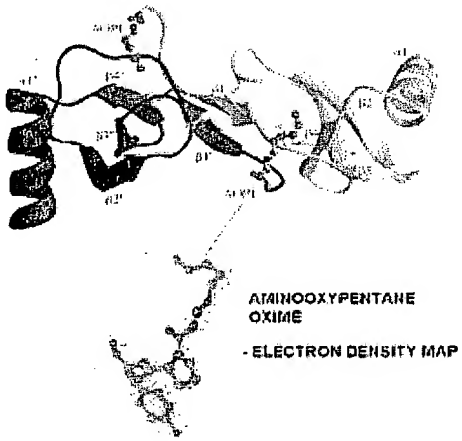


Figure 12 Crystal structure of AOP-RANTES (79). This chemically modified chemokine protein was prepared by total synthesis, using native chemical ligation. X-ray diffraction was used to determine the structure to 1.6-Å resolution. (*Top*) Ribbon structure of the crystalline dimer. (*Bottom*) $2F_o$-F_c electron density map corresponding to the unnatural aminooxypentane oxime moiety.

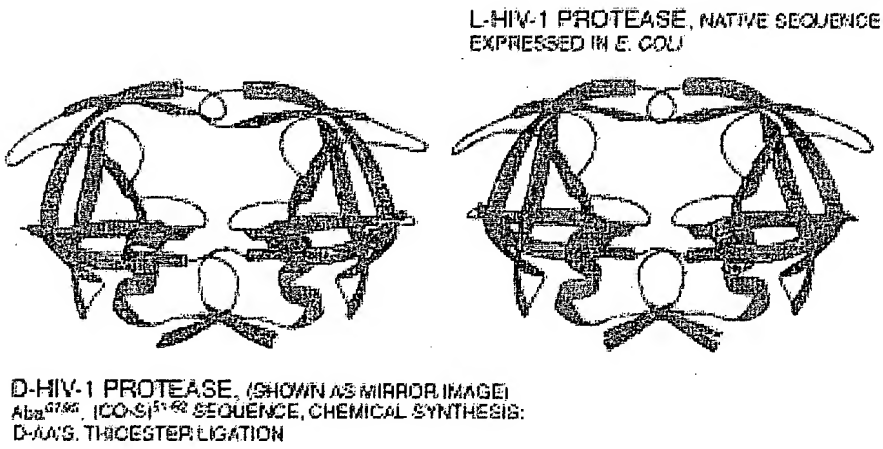


Figure 13 Three-dimensional crystal structure of the mirror image protein molecule, D-HIV-1 protease (20). The protein was prepared by thioester-forming chemical ligation of peptide segments synthesized with D-amino acids (52). (*Left*) Molscript representation of the ligated chemically synthesized D-protein molecule, displayed as the L-form for comparison purposes (20). (*Right*) Molscript representation of recombinant L-HIV-1 protease, prepared in *E. coli* (81a). The close similarity of the folded structures of the synthetic ligated D-protein and the recombinant L-protein is clearly evident.

and hexamers (82), as well as proteins containing two (57) and even three (82a, TM Hackeng, JA Fernandez, PE Dawson, SBH Kent, JH Griffin, submitted for publication) domains in a single polypeptide chain. The successful syntheses of such proteins suggests that this ability to accurately fold synthetic polypeptide chains may hold true both for single domains and for more complex proteins.

Certainly, in vitro folding, in which the system contains only a single homogeneous polypeptide of defined covalent structure, is utterly distinct from and much simpler than the situation in vivo, in which the complex intracellular environment contains multiple interacting protein species at high local concentrations. In consequence of this complexity, it has recently been found that the cell possesses a sophisticated "chaperone" apparatus that is involved in protein folding in vivo (83).

In chemical protein synthesis, the folded protein molecule is formed only at the final stages of production, under carefully controlled laboratory conditions. Control of the folding process can be particularly important in the production of proteins that are toxic to the cell, such as proteolytic enzymes. Using chemistry, it is possible to keep the polypeptide unfolded and inactive until after ligation and purification, when folding can be carried out in the presence of an inhibitor. This control over enzymatic activity was one of the key features of the success of chemical protein synthesis in the early work on the HIV-1 protease (20).

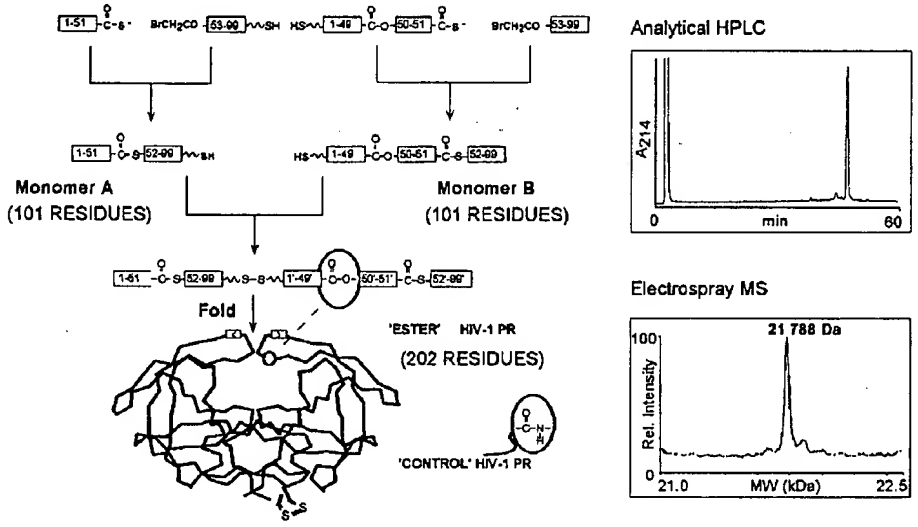


Figure 14 Total chemical synthesis of a multidomain protein: a tethered-dimer form of the HIV-1 protease. (*Left*) Convergent ligation synthetic strategy (85). Thioester-forming ligation was used to make each of the 101-residue monomers, and then directed disulfide formation was used to make the 202-residue synthetic dimer. The —NH— moiety of Ile50, in one monomer only, was substituted for by an —O— by incorporation of an ester at that position in chemical synthesis of the peptide segment (20). (*Right*) Analytical data. (*Top right*). HPLC, showing high-purity product and autolysis fragments consistent with the full enzymatic activity of the backbone-engineered 22-kDa protein. (*Bottom right*) Electrospray mass spectrometry data showing the high purity and correct observed mass of the synthetic construct.

CASE STUDIES IN THE APPLICATION OF CHEMICAL PROTEIN SYNTHESIS

Noncoded Amino Acids

Using chemistry to make proteins, it is straightforward to introduce an almost unlimited range of "unnatural" amino acids at any specific site(s) in a protein molecule. A classic example is Low's investigation of the "second shell" effects on the redox potential of an iron-sulfur protein, by systematic substitution of non-coded amino acids (84). Demonstrating the power of the chemical protein synthesis method, large amounts of each protein analog were made, purified, and fully characterized. Another example is the incorporation of a thienyl-Ala in place of a His residue in the enzyme PLA2 to establish the critical function of an imidazole side-chain functionality in the action of that enzyme (65). With chemical synthesis, multiple substitutions can be readily made at any position of the polypeptide chain, enabling virtually unlimited combinations of number, type, and position of noncoded amino acids to be incorporated into a protein molecule (85). Such studies can be very informative as to the structural basis of protein function and are made

straightforward by the synthesis of native proteins by chemical ligation of unprotected synthetic peptide segments, yet they are extremely difficult or impossible by recombinant DNA methods.

Noncoded amino acids are frequently found in native proteins in vivo. These arise from specific post-translational enzymatic modification of coded amino acid residues. One common modification of this type is γ-carboxy-glutamic acid (Gla), found for example in the eponymous Gla domains in plasma proteins. This modified amino acid is not produced biosynthetically in bacteria or yeasts, which rules out simple expression, so chemical synthesis offers an attractive route to the preparation of proteins containing Gla domains.

An example is human plasma protein S, a 635-amino-acid (aa) plasma protein that acts as an anticoagulant cofactor. This multidomain protein consists of an N-terminal Gla domain that contains 11 Gla residues, which is followed by a thrombin-sensitive region, three epidermal growth factor domains, and a sex hormone-binding globulinlike region. A polypeptide construct containing the first three domains, Gla (1–46 aa)–thrombin-sensitive region (47–76 aa)–epidermal growth factor-1(77–116 aa), has been synthesized from three segments, using native chemical ligation (TM Hackeng, JA Fernandez, PE Dawson, SBH Kent, JH Griffin, submitted for publication). Folding of this polypeptide chain produced a three-domain protein, microProtein S, that displayed anticoagulant cofactor activity.

Precise Covalent Modification

The ability to prepare native proteins by total synthesis, using chemical ligation of unprotected peptide segments, provides a convenient and general route to site-specific modification of the protein molecule. The full range of synthetic peptide and peptidomimetic chemistry (86) is at the command of the researcher who wants to make precise and controlled changes in a protein's covalent structure. Such changes are not limited by the genetic code or by the strictures of the ribosomal machinery. With chemical synthesis, virtually any conceivable covalent modification can be introduced at will anywhere in the protein molecule. An early example of the utility of this approach was the total chemical synthesis of (BTD) HIV-1 protease (87), a protein in which the Gly-Gly sequence found in a β-turn in the native protein (20) was replaced by the sterically constrained bicyclic compound BTD, a rigid mimetic of type II′ β-turn geometry (88). The resulting enzyme showed full activity and a significantly enhanced thermostability (87).

Site-Specific Tagged Proteins

Chemical synthesis enables the specific labeling of a protein molecule at unique site(s). Such specific modification is less likely to perturb the structure or activity of the protein than uncontrolled reaction with labeling reagents that stochastically target all amino or other particular functional groups in the protein. Fluorescently tagged proteins are extremely useful tools for biology and drug discovery, and synthesis of native proteins by chemical ligation allows the facile incorporation of

fluorescent dye molecules at any desired position in a protein molecule. In a recent example, a fluorescent Trp analog was incorporated by chemical synthesis into the Ras-binding domain of the protein Raf (89). The binding properties of the native domain were maintained, and the unique fluorescent label permitted the study of extremely fast kinetics of protein-protein binding.

With chemical protein synthesis, it is possible to tune the fluorescent properties of the labeled protein to the task at hand. For example, dye chelator-labeled proteins have been made for time-resolved fluorescence studies, in which it is possible to largely eliminate background emission by use of suitably "time-gated" detection (C Hunter, G Kochendoerfer, 89a).

Finally, total chemical synthesis allows the ready introduction of affinity tags, such as biotin, at precise sites in the protein molecule, while preserving biological activity, again something that is straightforward with chemistry.

Backbone Engineering

Another intriguing example of site-specific modification of the protein molecule, enabled by chemical ligation, is the covalent modification of the polypeptide backbone itself. This type of modification is not readily achieved, if possible at all, by genetic-engineering means. For example, a functionally important peptide bond (i.e. backbone amide) in the HIV-1 protease molecule was site-specifically replaced by a thioester moiety in each monomer of the homodimeric protein molecule, to investigate the direct involvement of that specific peptide bond in the mechanism of action of the enzyme, as suggested by the X-ray crystallographic data (20, 56). This approach was extended to the construction by total chemical synthesis of a 22-kDa covalent tethered dimer of the HIV-1 protease (20, 85), in which only one monomer was site specifically backbone engineered (Figure 14). The results of these studies showed that the two flap regions of the homodimeric native HIV-1 protease molecule work analogously to the single flap moiety in the two-domain, single-polypeptide chain, cell-encoded aspartyl proteinases (90).

A similar backbone engineering approach, in which specific amide —NH— moieties were replaced by —O— atoms, has been used to investigate the contribution of individual backbone H bonds to protein-protein interactions (91). More recently, an engineered backbone structure was introduced into bovine pancreatic trypsin inhibitor, by replacing one Cys residue involved in forming a disulfide bond with an *N*(ethylmercaptan)Gly, to investigate the effects of such a substitution on the folding, activity, structure, and stability of the resulting protein molecule (92). In this novel protein analog, the side chain of the Cys residue has effectively been moved to the backbone amide N atom.

Protein Medicinal Chemistry

Synthetic access enables the systematic application of the principles of medicinal chemistry to the protein molecule itself. An example is the total chemical synthesis of the potent anti-HIV molecule AOP-RANTES (79) (Figure 12). This chemical

protein analog was used as a lead compound in a successful program, based on chemical ligation, to develop even more potent anti-HIV molecules (J Wilken, D Thompson, H Gaertner, O Hartley, R Fish, JM McDonnell, Q Xu, D Fushman, D Cowburn, N Heveker, J Picard, SBH Kent, R Offord, manuscript in preparation). The chemical protein analog NNY-RANTES, which resulted from the first phase of this program, is >30-fold more effective as an anti-HIV compound and has been shown to prevent HIV infection at low nanomolar concentrations in the huPBL-SCID mouse model for acuired immune deficiency syndrome (93). NNY-RANTES is the most potent known anti-HIV compound. It is believed to work by inhibiting receptor recycling (94), thus clearing CCR5 from the surface of peripheral blood cells, a mechanism distinct from current clinical therapies for acquired immune deficiency syndrome.

Rapid Access to Functional Gene Products

In the past few years, an important new application has emerged for chemical protein synthesis—to enable rapid access to functional wild-type protein molecules directly from gene sequence data (Figure 15). Success of the genome projects has resulted in the discovery of >100,000 new proteins (1). However, these newly discovered molecules are known only as predicted open reading frames in genome sequence databases—the biomedical researcher rarely has access even to the cDNA clone corresponding to a particular gene, let alone the protein itself. For example, the recent elucidation of the complete DNA sequence of the genome of *Caenorhabditis elegans* resulted in the identification of 19,090 open reading frames encoding ~7.5 million amino acid residues of polypeptide sequence (95)! The probable roles of many of these predicted proteins can be tentatively assigned by analogy to proteins of known function, using bioinformatics. Nevertheless, the precise biochemical properties of a mature gene product can only be assessed at the level of the protein molecule itself.

Synthesis of native proteins by chemical ligation of unprotected peptides can provide access in a matter of days to large (10^{+}mg) amounts of functional protein molecules of exquisite homogeneity, based directly on gene sequence data. Secretory proteins, which are generally small and rich in Cys residues, are particularly suited to facile preparation by native chemical ligation. As described above, over the past 3 years, >300 proteins and protein analogs have been prepared by this method (78). These synthetic proteins have been used in a wide range of biomedical research investigations, resulting not only in the definition of gene function but frequently in the elucidation of novel biology (96).

Structural Biology

Facile access to the large (i.e. multiple tens of milligram) amounts of high-purity preparations produced by chemical protein synthesis can be of great utility for studies of protein structure by NMR spectroscopy and by X-ray crystallography. New methods for NMR spectroscopy have considerably enhanced the speed with which the structure of small (i.e. <200-aa-residue) proteins can be determined.

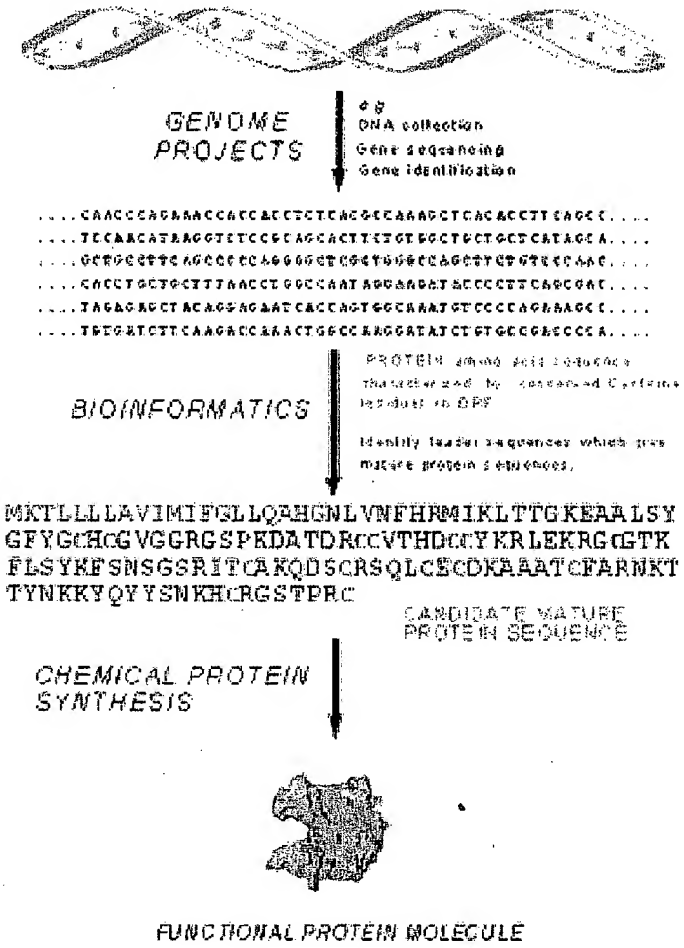


Figure 15 From gene sequence direct to functional protein, using chemical protein synthesis.

Obtaining sufficient (i.e. >10-mg) amounts of correctly folded proteins is now often the limiting step in structure determination. In a number of instances, total synthesis by chemical ligation methods has provided rapid access to high-purity protein samples in amounts useful for NMR studies (97, 98).

A recent case study of the determination by NMR of the novel structure of a chemically synthesized protein is the C-terminal Cys-rich domain of the "agouti-related" protein (80), a natural antagonist of the melanocortin-4 receptor involved in the control of human feeding behavior. In addition to small protein domains, chemical ligation approaches have contributed to the analysis of large proteins, using NMR techniques. Muir and coworkers have made use of recombinant

expression of N-terminal cysteine and thioester polypeptides (folded as domains) to label individual domains within multidomain proteins (66). By reducing spectral complexity, this approach promises to greatly simplify the NMR analysis of proteins that are >200 amino acids in size.

Chemistry also enables the precise site-specific introduction of NMR probe nuclei into the protein molecule. Thus, for the HIV-1 protease, the single γ-C atom of the active-site Asp side-chain carboxylate in each protein subunit was uniquely ^{13}C-labeled (99). NMR measurements in the presence and absence of inhibitor showed distinctive chemical shifts as a function of pH. It was possible to define the protonation state of the enzyme's catalytic apparatus and, from the unusual and dramatic chemical shifts observed, to deduce the molecular basis of the enhanced nucleophilicity of one of the two Asp side chains at the active site. It is this "super nucleophilicity" that is the defining feature of aspartyl proteinases as a class (100). This ability to precisely define at the level of a single functional group the unique molecular basis of enzymatic properties demonstrates the power of chemistry applied directly to the protein molecule itself.

Similarly, new X-ray crystallography methods have accelerated the pace of protein structure determination. In increasing instances, protein synthesis by chemical ligation has been used in conjunction with X-ray crystallography to determine the structures of novel proteins. Examples include, the chemokine SDF-1α (81), the chemical protein analog AOP-RANTES (79; Figure 12), and the mirror-image enzyme molecule D-HIV-1 protease (20; Figure 13).

Another important application of chemical protein synthesis is in the emerging genomic structural biology programs, which are aimed at the determination of the three-dimensional molecular structures of representative examples from all classes of proteins encoded in a particular genome (101). Such high-throughput structure determination will require access to great numbers of proteins in high purity and large amount. In addition, incorporation into the protein molecule of seleno-methionine residues is essential to also provide direct phase information from anomalous X-ray scattering on the same protein sample. Chemical protein synthesis by the methods described here is well suited to provide the proteins needed for genomic structural biology. A pilot study has been successfully completed in which the viral chemokine vMIP-II was prepared in [Se]Met-containing form and used for structure determination by both ^{1}H-NMR (98) and X-ray crystallography (E Lolis, submitted for publication).

CURRENT DEVELOPMENTS

Expressed Protein Ligation

From its inception, the native chemical ligation method was also envisioned for use with peptides that are produced by recombinant means (62). There are now multiple examples of the chemical ligation of recombinant peptides. These alternate

sources for suitably functionalized peptides have extended the applicability of the native chemical ligation method to include the world of peptides and domains that can be successfully produced by recombinant-DNA-based expression methods.

N-terminal cysteine recombinant peptides can be generated either by proteolytic cleavage next to a cysteine residue (102) or by an intein-based approach (103). These recombinant products can be reacted with synthetic peptide-thioesters to generate native polypeptides of hybrid biological and chemical origin. More recently, intein-based protein expression vectors have been adapted to generate polypeptide thioesters by recombinant means for use in native chemical ligation (104, 105). Interception of the partly rearranged splicing intermediate by a suitable thiol generates a recombinant peptide-thioester (Figure 16). These peptide-thioester segments can be reacted by native chemical ligation with a synthetic N-terminal Cys peptide to generate native polypeptides of hybrid chemical and biological origin (104–107). With the approaches described above, both the peptide-thioester and the N-terminal Cys peptide can be of recombinant origin. This permits the use of native chemical ligation for the mixing and matching of recombinant polypeptide segments in vitro (66).

Use of recombinant methods to generate the necessary peptide-thioester segments thus permits even molecular biologists who are not skilled in chemistry to use the native chemical ligation technique (106, 107). This "expressed protein ligation"[3] can be expected to lead to widespread use of the native chemical ligation method in biological research laboratories (109, 110).

Solid-Phase Protein Synthesis

The principles of polymer-supported organic synthesis (13, 19, 111) have been applied to the chemical ligation of unprotected peptide segments in aqueous solution [(112); Figure 17]. In solid-phase chemical ligation, unprotected peptide segments of 35–50 amino acids (i.e. ~5 kDa each) are used as building blocks to assemble the target polymer-bound polypeptide by consecutive ligation on a water-compatible polymer support. Strategies for segment condensation in both the N-to-C and C-to-N directions have been used successfully for solid-phase protein synthesis (112) and alternative linker chemistries developed (112a).

Target molecules have been constructed from as many as eight peptide segments by solid-phase chemical ligation [e.g. the polypeptide of the tissue plasminogen activator catalytic domain; M_w 25,000 (W Lu, unpublished data), and the polypeptide chain of the enzyme secretory PLA2 GV has beenassembled in a single day

[3]Sometimes erroneously referred to as "intein-mediated ligation" (106). It is important to note that, where incipient splicing of a defective intein is simply used as a way of generating a (recombinant) peptide-thioester, the ligation itself is not intein mediated; rather, the ligation reaction is standard native chemical ligation of two unprotected peptide segments (62). For an example of true intein-mediated ligation, see Reference 108.

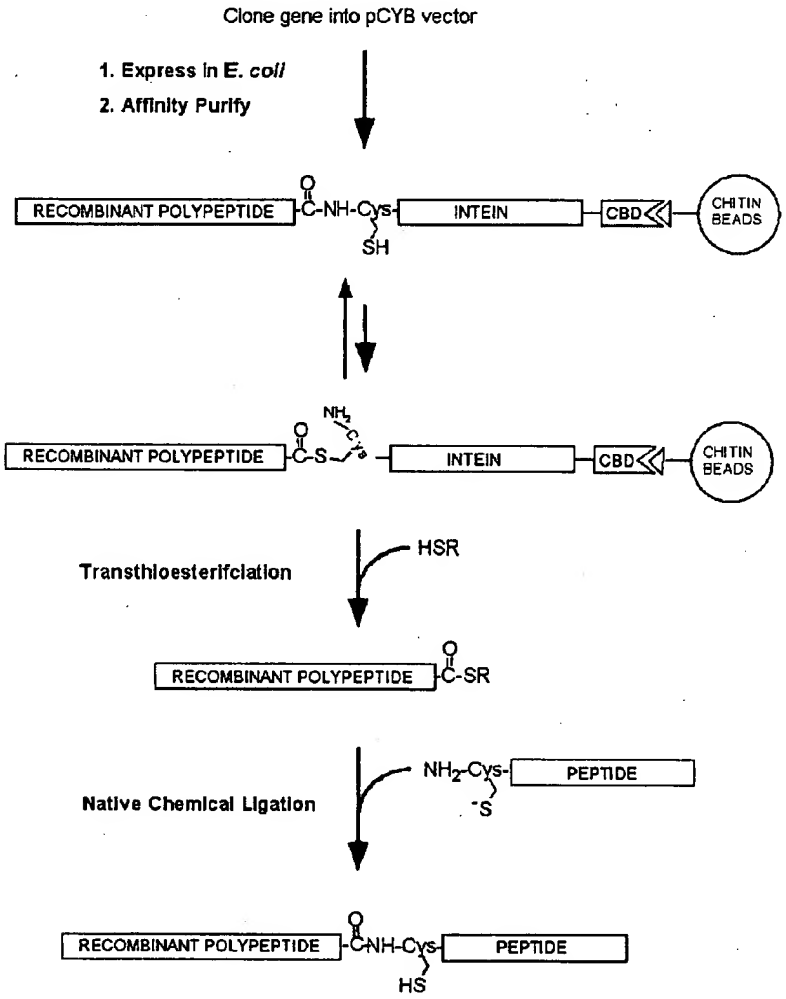


Figure 16 Expressed protein ligation (104). This process uses intein-mediated (73) preparation of a recombinant peptide-$^{\alpha}$thioester, which is then reacted with a Cys-peptide segment by native chemical ligation to prepare the desired product.

from four peptide segments (112). It can be anticipated that solid-phase chemical ligation will provide a practical chemical route to proteins that contain several hundred amino acids (Figure 18).

Membrane Proteins

An important aspect of the study of proteins which have been predicted from gene sequence data is the integral membrane class of proteins. Computer-aided analysis of the predicted open reading frames from a number of completely sequenced

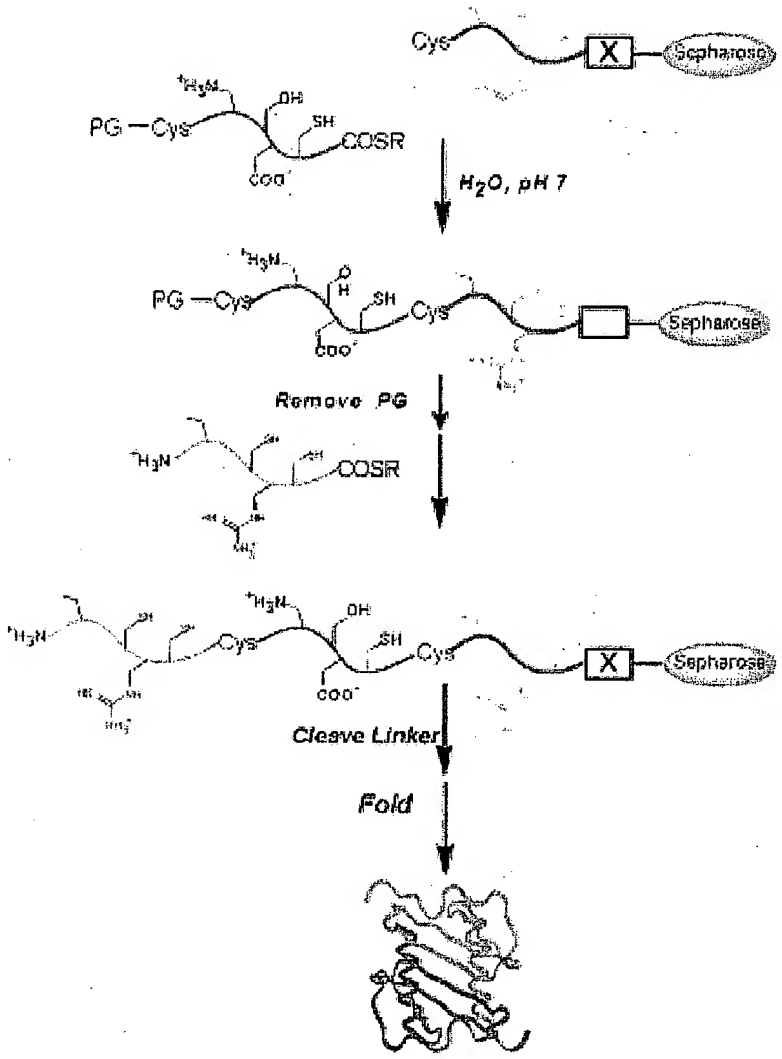


Figure 17 Solid-phase chemical ligation (112). Native chemical ligation of unprotected peptide segments and the principles of polymer-supported synthetic organic chemistry (13, 19, 111) are applied to solid-phase protein synthesis. In the example shown, the C-terminal segment of the target polypeptide is attached by a cleavable linker to a water-compatible support. The next segment as a peptide-$^{\alpha}$thioester is reacted by native chemical ligation, to give the polymer-bound ligation product. After removal of the Cys-protecting group (*PG*), successive rounds of ligation can be carried out to give the polymer-bound target polypeptide. After cleavage from the polymer support, the product is purified and folded to give the target protein molecule.

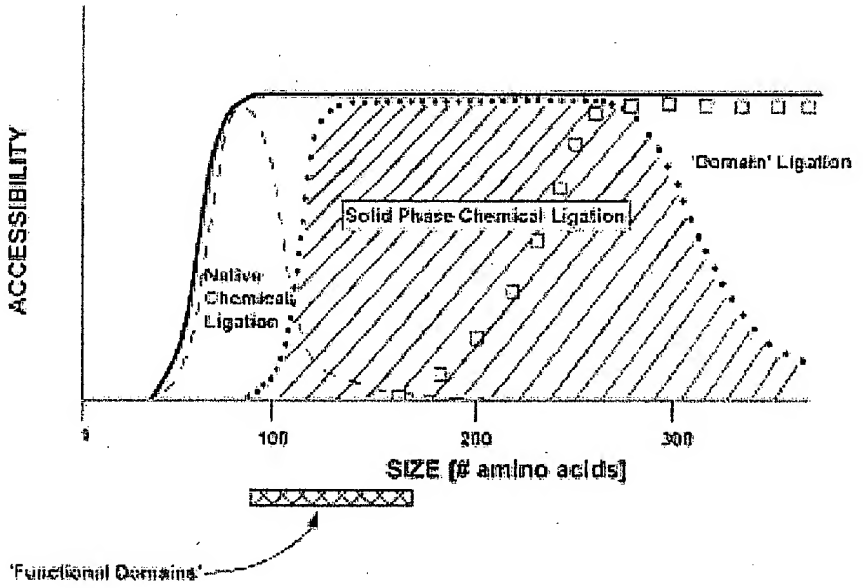


Figure 18 Size of synthetic polypeptides accessible by chemical ligation.

genomes has suggested that 20%–30% of all proteins contain membrane-spanning polypeptide sequences in the mature form of the molecule (113). Such integral membrane proteins mediate many processes in the cell, including signal transduction, ion transport, and active transport of macromolecules to name a few significant biological activities, and are thus important objectives for biomedical research. Yet integral membrane proteins are difficult to express at high levels by recombinant-DNA-based methods and have proven hard to isolate in homogeneous form in chemically defined media (114).

It is interesting that Kochendoerfer et al (115) have shown that integral membrane proteins can be synthesized in large amounts by the chemical ligation of unprotected peptide segments and isolated in high purity in media of defined chemical composition. An example is the total synthesis of the 11-kDa proton channel M2 protein of influenza A virus, which forms a tetrameric ion channel (115; Figure 19). The M2 protein had previously proven refractory to multiple attempts at expression by recombinant-DNA-based methods (W Degrado, personal communication), but was readily obtained by chemical ligation of unprotected synthetic peptides.

Glycoprotein Synthesis

Recently, the Bertozzi laboratory (116) reported the first total synthesis of a glycoprotein, using native chemical ligation in conjunction with innovative methods for the synthesis of glycopeptide-$^{\alpha}$thioesters. One of the most important applications

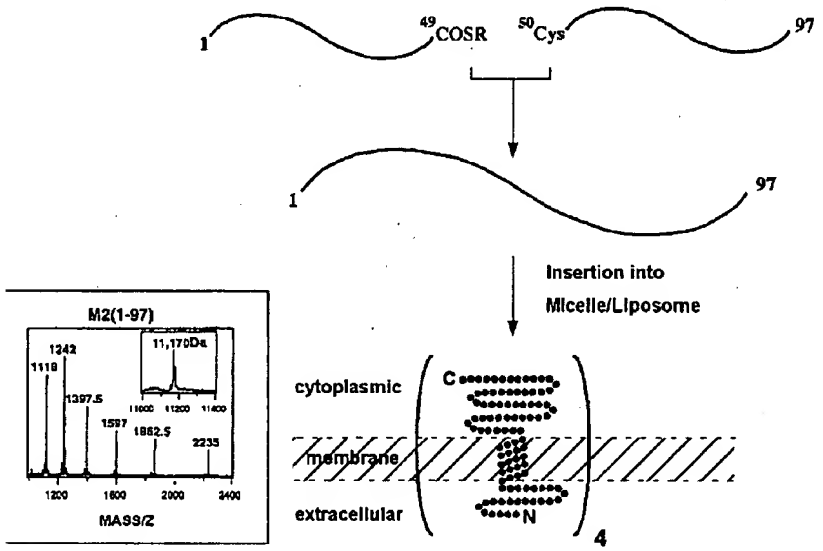


Figure 19 Chemical synthesis of an integral membrane protein (115). The 97-residue polypeptide chain of the influenza M2 protein was prepared by native chemical ligation and folded to form the active tetrameric form. (*Insert*) Electrospray mass spectrometric data showing the desired product, mass 11,170 Da.

of chemical protein synthesis will be the systematic preparation of glycoforms of gylcosylated proteins as homogeneous molecular species of defined covalent structure, to establish the role of the carbohydrate moiety in the biological function of the glycoprotein. In the near future, we can expect to see an increasing number of examples of this important capability made possible by native chemical ligation (62) and by other chemoselective reactions (117).

FUTURE DEVELOPMENTS

Ligation Sites

In its current form, native ligation chemistry uses a Cys residue at the site of formation of the new peptide bond joining two unprotected peptide segments. This means that, for a protein to be accessible by native chemical ligation, there must be no region in the polypeptide chain > 50–60 aa residues without at least 1 Cys residue. Although the requirement for a Cys at the ligation site may superficially be viewed as a stringent limitation of the method, it is actually less restrictive than it at first seems. There are hundreds of protein families with interesting biological activities, encompassing many thousands of protein molecules that have native Cys residues located in positions compatible with direct application of native chemical ligation (118).

In actual practice virtually any protein molecule can be made by native chemical ligation. For proteins with no suitable Cys ligation sites in the natural sequence, it is possible to simply put a Cys wherever one is needed for ligation, usually without deleterious effects on function (63, 67; see Figure 9). The work of Muir and coworkers is illustrative of this expedient but effective approach (66, 104, 110). Their chemical ligation of recombinantly expressed polypeptide-$^{\alpha}$thioesters to synthetic peptides has typically made use of an arbitrarily introduced Cys residue at the desired ligation site, with no deleterious effects. Also, biological researchers frequently insert Cys residues into a polypeptide chain to investigate the structure-function relationships in a protein molecule (119) or as a site for the introduction of a spectroscopic probe, such as an electron spin resonance label (120). This proven utility of arbitrarily introduced Cys residues provides considerable flexibility in synthetic design for the preparation of functional protein molecules by native chemical ligation at Cys.

Additionally, it would be desirable to have the option to use thioester-mediated chemical ligation at residues other than Cys. A prototype procedure for the use of an auxiliary-functional-group approach to native amide-forming, thioester-mediated chemical ligation has been reported (121). This work defined the principles of an effective approach to ligation at non-Cys residues, but the chemistry used had to be refined and extended because severe shortcomings were observed in the original investigation, as revealed by studies in model systems (121). In this respect, recently reported work from the Dawson laboratory at The Scripps Research Institute may represent a more effective chemistry for ligation at residues other than Cys (121a), using the same auxiliary-functional-group approach.

Size of Protein Targets

To date, it has proved possible to make every protein that has been attempted by the chemical ligation of unprotected peptide segments in aqueous solution, even integral membrane proteins. However, some targets are significantly more work than others—especially if there are multiple intermediate ligation products to handle. The recently developed solid-phase protein synthesis method (see above), using polymer-supported chemical ligation (112), provides a very effective means for the ready isolation of these intermediate products and will significantly simplify syntheses requiring ligation of multiple segments.

The work of our own and others' laboratories, including the laboratories of Offord (University of Geneva, Switzerland) and Muir (Rockefeller University, New York, NY), has failed to show any inherent size limitations for application of the chemical ligation method, up to several-hundred kilodaltons in the latter case (122). Folding of chemically synthesized polypeptide chains to form native proteins, in which significant problems might have been anticipated, is usually straightforward for the domain size proteins made to date. Folding of complex multidomain proteins may or may not be as straightforward. In any event, unlike expression systems, chemical ligation allows the option of constructing complex

proteins by separately folding each domain and then stitching the folded domains together (123).

Chemical Synthesis of Peptide Segments

Virtually any target protein can be prepared by total chemical synthesis, provided that a suitable set of high-purity peptide-thioester segments is available. Ironically, for many researchers the most challenging aspect of applying the chemical ligation method to proteins is making the peptide segments. To date, the principal constraint on widespread application of the native ligation method has been the lack of methods for the facile chemical synthesis of unprotected peptide-$^{\alpha}$thioester segments. Fortunately, there is an abundance of expertise available for the chemical synthesis of peptides (86). The need to make large numbers of analogs of thousands of native proteins by chemical ligation, and hence to prepare many tens-of-thousands of peptide segments, provides an unprecedented impetus for the development of efficient methods of peptide synthesis. We can look forward with confidence to the development of radically improved methods for the rapid, cost-effective preparation of large numbers of unprotected peptide-thioester segments for use in chemical protein synthesis (124).

SUMMARY AND CONCLUSIONS

Total synthesis by the chemical ligation of unprotected peptide segments can now provide general access to native proteins of $\leq$30 kDa (Figure 18) in size. This size range encompasses the structural and functional domains that are the modular building blocks of function in the protein world, from enzymes to receptors, from signal transduction adaptor molecules to large multisubunit protein assemblies. A wide range of different proteins has already been synthesized, leading to novel biology, new three-dimensional structures, and new insights into the molecular basis of protein function. In addition, it has already been demonstrated that it is possible to stitch together, by chemical ligation folded protein domains of any size, promising general access to the world of proteins.

Perhaps the most significant future application of chemistry to proteins will be in the creation, at will, of stable post-translational modified forms of protein molecules as homogeneous entities of precise covalent structure. This will enable the dissection at the level of the protein molecule of important biochemistry, such as the intracellular signal transduction pathways. It will also enable the systematic creation of new classes of protein therapeutics with enhanced properties.

The stage is now set for the application of the tools of chemistry to the entire universe of proteins. Truly, as Edward O. Wilson has remarked. "Where nucleic acids are the codes, proteins are the *substance* of life" (125). It is no exaggeration to say that understanding the molecular basis of protein action is one of the most important challenges of our era. The ability to apply chemistry to the study of

proteins, provided by the synthetic tools described in this article, will play an important part in addressing this challenge and will have a revolutionary impact on our understanding of gene function expressed through the medium of the protein molecule.

ACKNOWLEDGMENTS

We thank our colleague Dr. Manuel Baca for his critical reading of this chapter and for the many useful suggestions that he made. This article is a perspective on the synthesis of proteins by chemical means and, at the request of the Editors, emphasizes the work of the authors' own laboratories. Every attempt has been made to cite the original literature for all of the results described. The successes of chemical protein synthesis are due solely to the hard work of our many talented colleagues. The shortcomings of this review are entirely the responsibility of the authors.

LITERATURE CITED

1. Strasberg RL, Feingold EA, Klausner RD, Collins FS. 1999. *Science* 286:455–57
2. Dyson F. 1998. *The Red Herring Mag., June.* http://www.herring.com/mag/issue55/think.html
3. Matthews BW. 1993. *Annu. Rev. Biochem.* 62:139–60
4. Smith M. 1994. *Angew. Chem. Int. Ed. Engl.* 33:1214–21
5. Cleland JL, Craik CS. 1996. *Protein Engineering: Principles and Practice.* New York: Wiley & Sons. 518 pp.
6. Hecht SM. 1992. *Acc. Chem. Res.* 25:545–52
7. Mendel D, Cornish VW, Schultz PG. 1995. *Annu. Rev. Biophys. Biomol. Struct.* 24:435–62
8. Barrett JE, Lucero CM, Schultz PG. 1999. *J. Am. Chem. Soc.* 121:7965–66
9. Cornish VW, Mendel D, Schultz PG. 1995. *Angew. Chem. Int. Ed. Engl.* 34:621–33
10. Fruton JS. 1992. A *Skeptical Biochemist*, p. 137. Cambridge, MA: Harvard Univ. Press.
11. Bergmann M, Zervas L. 1932. *Chem. Berichte* 65:1192–201
12. Sheehan JC, Hess GP. 1956. *J. Am. Chem. Soc.* 77:1067–68
13. Merrifield RB. 1986. *Science* 232:341–47
14. Berman AL, Kolker E, Trifonov EN. 1994. *Proc. Natl. Acad. Sci. USA* 91:4044–47
15. Gerstein M. 1998. *Fold. Des.* 3:497–512
16. Xu D, Nussinov R. 1997. *Fold. Des.* 3:11–17
17. Doolittle RF, Bork P. 1993. *Sci. Am.* 268:50–56
18. Doolittle RF. 1995. *Annu. Rev. Biochem.* 64:287–314
19. Kent SBH. 1988. *Annu. Rev. Biochem.* 57:957–84
20. Miller M, Baca M, Rao JKM, Kent S. 1998. *J. Mol. Struct.* (THEOCHEM) 423:137–52
21. Ratner L. 1993. *Perspect. Drug Dis. Des.* 1:3–22
22. Schneider J, Kent SBH. 1988. *Cell* 54: 363–68
23. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, et al. 1989. *Science* 245:616–21
24. Navia M, Fitzgerald PM, McKeever BM, Leu CT, Heimbach JC, et al. 1989. *Nature* 337:615–20

25. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, et al. 1989. *Science* 246:1149–52
26. Swain AL, Miller MM, Green J, Rich DH, Schneider J, et al. 1990. *Proc. Natl. Acad. Sci. USA* 87:8805–9
27. Jaskolski M, Tomasselli AG, Sawyer TK, Staples DG, Heinrikson RL, et al. 1991. *Biochemistry* 30:1600–9
28. Lam PYS, Jadhav PK, Eyermann CJ, Hodge CN, Ru Y, et al. 1994. *Science* 263:380–84
29. Coffin J.1995. *Science* 267:483–89
30. Sieber P, Eisler K, Kamber B, Riniker B, Rittel W, et al. 1978. *Hoppe-Selyler's Z. Physiol. Chem.* 359:113–23
31. Akaji K, Fujii N, Yajima H, Hayashi K, Mizuta K, et al. 1985. *Chem. Pharm. Bull.* 33:184–201
32. Zawadzke LE, Berg JM. 1993. *Proteins* 16:301–5
33. Inui T, Bodi J, Kubo S, Hishio H, Kimura T, et al. 1996. *J. Pept. Sci.* 2:28–39
34. Hojo H, Kwon Y, Kakuta Y, Tsuda S, Tanaka I, et al. 1993. *Bull. Chem. Soc. Jpn.* 66:2700–6
35. Woo DD-L, Clarke-Lewis I, Chait BT, Kent SBH. 1989. *Protein Eng.* 3:29–37
36. Sakakibara S. 1995. *Biopolymers* 37:17–28
37. Kiyam S, Fujii N, Yajima H, Moriga M, Takagi A. 1984. *Int. J. Pept. Protein Res.* 23:174–86
38. Gatos D, Athanassopoulos P, Tzavara C, Barlos K. 1999. In *Peptides 1998*, ed. S Bajusz, F Hudecz, pp. 146–48. Budapest, Hungary: Akademiai Kiado
39. Clark-Lewis I, Kent SBH. 1989. In *The Use of HPLC in Protein Purification and Characterization*, ed. AR Kerlavage, pp. 43–79. New York: Liss
40. Chait BT, Kent SBH. 1992. *Science* 257:1885–94
41. Blake J, Li CH. 1981. *Proc. Natl. Acad. Sci. USA* 78:4055–58
42. Yamashiro D, Li CH. 1988. *Int. J. Pept. Protein Res.* 31:322–34
43. Kemp DS, Carey RI. 1993. *J. Org. Chem.* 58:2216–22
44. Hojo H, Aimoto S. 1991. *Bull. Chem. Soc. Jpn.* 64:111–17
45. Chang TK, Jackson DY, Burnier JP, Wells JA. 1994. *Proc. Natl. Acad. Sci. USA* 91: 12544–48
46. Braisted AC, Judice JK, Wells JA. 1997. *Methods Enzymol.* 289:298–313
47. Jackson DY, Burnier J, Quan G, Stanley M, Tom J, Wells JA. 1994. *Science* 266:243–47
48. Stephen BH, Kent D, Alewood P, Alewood M, Baca A, et al. 1992. In *Innovation and Perspectives in Solid Phase Synthesis*, ed. R Epton, pp. 1–22. Andover, UK: Intercept
49. Vogel AI. 1989. *Textbook of Practical Organic Chemistry*, p. 13. Reading, MA: Addison-Wesley
50. Schnolzer M, Kent SBH. 1992. *Science* 256:221–25
51. Muir TW, Kent SBH. 1993. *Curr. Opin. Biotechnol.* 4:420–27
52. deLisle-Milton R, Milton SCF, Schnolzer M, Kent SBH. 1993. In *Techniques in Protein Chemistry IV*, ed. RH Angeletti, pp. 257–67. New York: Academic
53. Rose K. 1994. *J. Am. Chem. Soc.* 116: 30–33
54. Dawson PE, Kent SBH. 1993. *J. Am. Chem. Soc.* 115:7263–66
55. Rau HK, Haehnel W. 1998. *J. Am. Chem. Soc.* 120:468–76
56. Baca M, Kent SBH.1993. *Proc. Natl. Acad. Sci. USA* 90:11638–42
57. Canne LE, Ferré-D'Amare AR, Burley SK, Kent SBH. 1995. *J. Am. Chem. Soc.* 117:2998–3007
58. Muir TW, Williams MJ, Ginsberg MH, Kent SBH. 1994. *Biochemistry* 33:7701–8
59. Englebretsen DR, Garnham BG, Bergman DA, Alewood PF. 1995. *Tetrahedron Lett.* 36:8871–74
60. Liu CF, Tam JP. 1994. *J. Am. Chem. Soc.* 116:4149–53

61. Liu CF, Rao C, Tam JP. 1996. *J. Am. Chem. Soc.* 118:307–12
62. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH. 1994. *Science* 266:776–79
63. Dawson PE, Churchill MJ, Ghadiri MR, Kent SBH. 1997. *J. Am. Chem. Soc.* 119:4325–29
64. Hackeng TM, Dawson PE, Griffin JH, Kent SBH. 1997. *Proc. Natl. Acad. Sci. USA* 94:7845–50
65. Hackeng TM, Griffin JH, Dawson PE. 1999. *Proc. Natl. Acad. Sci. USA* 96:10068–73
66. Xu R, Ayers B, Cowburn D, Muir TW. 1999. *Proc. Natl. Acad. Sci. USA* 96:388–93
67. Lu W, Randal M, Kossiakoff A, Kent SBH. 1999. *Chem. Biol.* 6:419–27
68. Lu WY, Qasim MA, Kent SBH. 1996. *J. Am. Chem. Soc.* 118:8518–23
69. Storer AC, Menard R. 1994. *Methods Enzymol.* 244:486–500
70. Scheffner M, Smith S, Jentsch S. 1998. In *Ubiquitin and the Biology of the Cell*, ed. J-M Peters, JR Harris, D Finley, pp. 65–98. New York: Plenum
71. Stachelhaus T, Mootz HD, Marahiel MA. 1999. *Chem. Biol.* 6:493–505
72. Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. 1989. In *Molecular Biology of the Cell*, pp. 1035–36. London: Garland
73. Paulus H. 2000. *Annu. Rev. Biochem.* 69:447–96.
74. Xu M-Q, Perler FB. 1996. *EMBO J.* 15:5146–53
75. Tam JP, Yu Q. 1998. *Biopolymers* 46:319–27
76. Beligere GS, Dawson PE. 1999. *J. Am. Chem. Soc.* 121:6332–33
77. Wallace CJA. 1995. *Curr. Opin. Biotechnol.* 6:403–10
78. Wilken J, Kent SBH. 1998. *Curr. Opin. Biotechnol.* 9:412–26
79. Wilken J, Hoover D, Thompson DA, Barlow PN, McSparron H, et al. 1999. *Chem. Biol.* 6:43–51
80. Bolin KA, Anderson DJ, Trulson JA, Gantz I, Thompson DA, et al. 1999. *FEBS Lett.* 451:125–31
81. Dealwis C, Fernandez EJ, Thompson DA, Simon RJ, Siani MA, Lolis E. 1998. *Proc. Natl. Acad. Sci. USA* 95:6941–46
81a. Fitzgerald PMD, McKeever BM, VanMiddlesworth JF, Springer JP, James P, et al. 1990. *J. Biol. Chem.* 265(24):14209–19
82. Fitzgerald MC, Chernushevich I, Standing KG, Kent SBH, Whitman CP. 1995. *J. Am. Chem. Soc.* 117:11075–80
82a. Beligere GS, Dawson PE. 1999. *Biopolymers* 51:363–69
83. Sigler PB, Xu Z, Rye HS, Burston SB, Fenton WA, Horwich AL. 1998. *Annu. Rev. Biochem.* 67:581–608
84. Low DW, Hill MG. 1998. *J. Am. Chem. Soc.* 120:11536–37
85. Baca M, Muir TW, Schnolzer M, Kent SBH. 1995. *J. Am. Chem. Soc.* 117:1881–87
86. Goodman M, Felix A, Moroda L, Toniolo C, eds. 2000. *Houben-Weyl: Methods of Organic Chemistry*, Vol. E 22: *Synthesis of Peptides and Peptidomimetics*, Stuttgart, Ger.: Thieme. In press
87. Baca M, Alewood PF, Kent SBH. 1993. *Protein Sci.* 2:1085–91
88. Nagai U, Sato K. 1985. *Tetrahedron Lett.* 26:647–50
89. Sydor JR, Herrmann C, Kent SBH, Goody RS, Engelhard M. 1999. *Proc. Natl. Acad. Sci. USA* 96:7865–70
89a. Kochendoerfer GG, Hunter CL. 1999. In *Peptides 1998*, ed. S. Bajusz, F. Hudecz, pp. 182–83. Budapest, Hungary: Akademiai Kiado
90. Davies DR. 1990. *Annu. Rev. Biophys. Biophys. Chem.* 19:189–215
91. Lu W, Qasim MA, Laskowski M Jr, Kent SBH. 1997. *Biochemistry* 36:673–79
92. Bark SJ, Kent SBH. 1999. *FEBS Lett.* 460:67–76
93. Mosier DE, Picchio GR, Gulizia RJ, Sabbe R, Poignard P, et al. 1999. *J. Virol.* 73:3544–50

94. Mack M, Luckow B, Nelson PJ, Cihak J, Simmons G, et al. 1998. *J. Exp. Med.* 187:1215–24
95. The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*. *Science* 262:2012–18
96. Campbell JH, Zlotnik A, Siani MA, Thompson DA, Butcher EC. 1998. *Science* 279:381–84
97. Lu W, Starovasnik MA, Kent SBH. 1998. *FEBS Lett.* 429:31–35
98. Shao W, Fernandez E, Wilken J, Thompson DA, Siani MA, et al. 1998. *FEBS Lett.* 441:77–82
99. Smith R, Brereton IM, Chai RY, Kent SBH. 1996. *Nat. Struct. Biol.* 3: 946–50
100. Hartsuck JA, Tang J. 1972. *J. Biol. Chem.* 247:2575–80
101. Pennisi E. 1998. *Science* 279:978–79
102. Erlanson DA, Chytil M, Verdine GL. 1996. *Chem. Biol.* 3:981–91
103. Perler FB, Xu MQ, Paulus H. 1997. *Curr. Opin. Chem. Biol.* 1:292–99
104. Muir TW, Sondhi D, Cole PA. 1998. *Proc. Natl. Acad. Sci. USA* 95:6705–10
105. Kinsland C, Taylor SV, Kelleher NL, Mclafferty FW, Begley TP. 1998. *Protein Sci.* 7:1839–42
106. Roy RS, Allen O, Walsh CT. 1999. *Chem. Biol.* 6:789–99
107. Evans TC, Benner J, Xu M-Q. 1998. *Protein Sci.* 7:2256–64
108. Otomo T, Ito N, Kyogoku Y, Yamazaki T. 1999. *Biochemistry* 38(49):16040–44
109. Peters R, Sikorski R. 1998. *Science* 281:367–68
110. Cotton GJ, Muir TW. 1999. *Chem. Biol.* 6:R247–56
111. Bunin BA. 1998. *The Combinatorial Index.* New York: Academic. 322 pp
112. Canne LE, Botti P, Simon RJ, Chen Y, Dennis EA, Kent SBH. 1999. *J. Am. Chem. Soc.* 121:8720–27
112a. Brik A, Keinan E, Dawson PE. 2000. *J. Org. Chem.* 65: In press
113. Wallin E, von Heijne G. 1998. *Protein Sci.* 7:1029–38
114. England PM, Zhang Y, Dougherty DA, Lester HA. 1999. *Cell* 96:89–98
115. Kochendoerfer GG, Salom D, Lear JD, Wilk-Orescan R, Kent SBH, DeGrado WF. 1999. *Biochemistry* 38:11905–13
116. Shin Y, Winans KA, Backes BJ, Kent SBH, Ellman JA, Bertozzi CR. 1999. *J. Am. Chem. Soc.* 121:11684–89
117. Lemieux GA, Bertozzi CR. 1998. *Trends Biotechnol.* 16:506–13
118. Tsujimura A, Yasojima K, Kuboki Y, Suzuki A, Ueno N, et al. 1995. *Biochem. Biophys. Res. Commun.* 214:432–39
119. Cai K, Klein-Seetharaman J, Farrens D, Zhang C, Altenbach C, et al. 1999. *Biochemistry* 38:7925–30
120. Langen R, Cai K, Altenbach C, Khorana HG, Hubbell WL. 1999. *Biochemistry* 38:7918–24
121. Canne LE, Bark SJ, Kent SBH. 1996. *J. Am. Chem. Soc.* 118:5891–96
121a. Offer J, Dawson PE. 2000. *Org. Lett.* 2:23–6
122. Severinov K, Muir TW. 1998. *J. Biol. Chem.* 273:16205–9
123. Fitzgerald MC, Kent SBH. 1998. In *Bioorganic Chemistry: Peptides and Proteins*, ed. SM Hecht, pp. 65–99. New York: Oxford Univ. Press
124. Ingenito R, Bianchi E, Fattori D, Pessi A. 1999. *J. Am. Chem. Soc.* 121:11369–74
125. Wilson EO. 1998. In *Consilience: the Unity of Knowledge*, pp. 99–100. New York: Knopf